

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 04:01:44 ; Search time 25622 Seconds
(without alignments)
11676.376 Million cell updates/sec

Title: US-09-242-772-116
Perfect score: 7313
Sequence: 1 ggcagcgatacactacaat.....tatgaataaaatctcgtgcc 7313

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.on.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	7313	100.0	7313	9	HSU65002	U65002 Human zinc
2	7312.6	100.0	7313	6	A69511	A69511 Sequence 1
C 3	6520	89.2	142102	9	AC107952	AC107952 Homo sapi
C 4	2821.6	38.6	182102	2	AC097274	AC097274 Mus muscu
C 5	2821.6	38.6	241148	10	AL807387	AL807387 Mouse DNA
C 6	2761.2	37.8	278377	2	AC129839	AC129839 Rattus no
C 7	2623.8	35.9	244955	2	AC123210	AC123210 Rattus no
C 8	1295.6	17.7	1575	10	AF057366	AF057366 Mus muscu
C 9	707.4	9.7	65748	2	AC090187	AC090187 Homo sapi
C 10	651.8	8.9	65748	2	AC090187	AC090187 Homo sapi
C 11	645.4	8.8	49357	2	AC107961	AC107961 Homo sapi
C 12	416	5.7	421	11	G25280	G25280 human STS E
C 13	392.8	5.4	1829	9	BC023655	BC023655 Homo sapi
C 14	392.8	5.4	2429	9	AK026936	AK026936 Homo sapi
C 15	392.8	5.4	4051	9	AF006005	AF006005 Homo sapi
C 16	392.8	5.4	5638	9	D83784	D83784 Human mRNA
C 17	391.2	5.3	5267	10	AB051854	AB051854 Mus muscu
C 18	383.8	5.2	550	9	AF221550	AF221550 Homo sapi
C 19	364.8	5.0	1437	10	AF181262	AF181262 Mus muscu
C 20	359.2	4.9	161478	9	AC025538	AC025538 Homo sapi
C 21	352.6	4.8	145414	9	HSA392M18	AL121897 Human DNA
C 22	352.6	4.8	200329	10	AC078911	AC078911 Mus muscu
C 23	352.6	4.8	208614	10	AL807380	AL807380 Mouse DNA
C 24	352	4.8	240510	2	AC134371	AC134371 Rattus no
C 25	351.6	4.8	2334	6	BD005470	BD005470 Nucleic a
C 26	351.6	4.8	2828	9	HSA635	AJ006354 Homo sapi
C 27	351.6	4.8	3215	9	AK096331	AK096331 Homo sapi
C 28	351.6	4.8	3663	9	HSA311395	AJ311395 Homo sapi
C 29	351	4.8	240336	2	AC094844	AC094844 Rattus no
C 30	350	4.8	2815	6	AX329760	AX329760 Sequence
C 31	350	4.8	3162	9	HSU72621	AK091707 Homo sapi
C 32	350	4.8	4833	9	HSU72621	U72621 Homo sapien
C 33	344.4	4.7	1874	5	AF186476	AF186476 Danio rer
C 34	333	4.6	376	6	AR274991	AR274991 Sequence
C 35	301.8	4.1	5028	10	RNU72620	U72620 Rattus norv
C 36	295.8	4.0	2350	10	AF324471	AF324471 Mus muscu
C 37	294.4	4.0	2771	10	MMZINCFFR	X95503 M.musculus
C 38	294.4	4.0	2790	6	BD005460	BD005460 Nucleic a
C 39	294.4	4.0	3732	10	MMZINCFFP	X95504 M.musculus
C 40	294.4	4.0	3975	10	AF147785	AF147785 Mus muscu
C 41	291.4	4.0	236442	2	AC128089	AC128089 Rattus no
C 42	279.2	3.8	2467	9	HSA303119	AJ303119 Homo sapi
C 43	279.2	3.8	137387	9	HSJ468K18	AL049844 Human DNA
C 44	277.6	3.8	2561	6	A69513	A69513 Sequence 3
C 45	277.6	3.8	2561	9	HSU81992	U81992 Homo sapien

ALIGNMENTS

RESULT 1
HSU65002
LOCUS
DEFINITION Human zinc finger protein PLAG1 mRNA, complete cds.
ACCESSION U65002
VERSION U65002.1 GI:1839159
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 7313)
AUTHORS Kas.K., Voz.M.L., Roijer.E., Astrom,A.K., Meyen,E., Stenman,G. and Van de Ven,W.J.
TITLE Promoter swapping between the genes for a novel zinc finger protein

and beta-catenin in pleiomorphic adenomas with t(3;8) (p21;q12)
translocations
Nat. Genet. 15 (2), 170-174 (1997)
97172974
PUBMED
9020842
REFERENCE
2 (bases 1 to 7313)
Kas, K., Voz, M.L., Roijer, E., Meyen, E., Stenman, G. and Van de
Ven, W.J.M.
Direct Submission
Submitted (24-JUL-1996) Lab for Molecular Oncology, Center of Human
Genetics - K.U.Leuven, Herestraat 49, Leuven 3000, Belgium
Genetics
Location/Qualifiers
1..7313
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8q12"
1..159
/number=1
160..264
/number=2
265..362
/number=3
363..722
/number=4
481..1983
/codon_start=1
/product="zinc finger protein PLAG1"
/protein_id="AAC50995.1"
/db_xref="GI:1839160"
/translation="MATVPIQDLSEVRDQKVPKSGKRGKTPKRNFPKPCOLCDKAFN
SVKLVHSYSHTEPKYKIQDCTAFVSKYUQRHWAHSPEKHKNCNCEKMFH
RKDLKHLHLPNKETPKCEGKNYKLGFKKHLAHLAHSGLTKVCLQYFE
STGVLEHLKSHAGSGGKVKKHCQCHDRFYRDXVRRHVVHTGRKDFLCQYC
AORFGKDLHTRHMKSHNOELLKVKTEPVDLPDFTCNVSPIDKDLPLVMSLPSE
LSPSTYALISLPKQPLKEIESYLMELQGVPPSSQDSQASSSKLGLDPOIGSLD
DGADLSLKSSTISIDPLNTPALDFSLNFPIPLNGPPNPLSVLSGLMSYSQESAH
SSVSQPLPTQDQLQDPANTIGLSLHSLSAAFSTSSLSSTTLPRFHOAFQ"
723..7313
/number=5
2287 a 1327 c 1357 g 2342 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 7313; DB 9; Length 7313;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 7313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCAGCGCAGCCGAGAGGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGCCTT 180

Db 1 GGCAGCGCAGCCGAGAGGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGCCTT 180

QY 61 GCGCGTCAGCCGAGAGGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGCCTT 120

Db 61 GCGCGTCAGCCGAGAGGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGCCTT 120

QY 121 GCGCGGAGCGGCGGAGGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGCCTT 180

Db 121 GCGCGGAGCGGCGGAGGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGCCTT 180

QY 181 GCGCGTATTGGGACCCAGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGCCTT 240

Db 181 GCGCGTATTGGGACCCAGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGCCTT 240

QY 241 CTAGAGTTTGGAGCAGGCGCTCAGATTGGCCAAAATAGGAAGGATGGATTCCACTCT 300

Db 241 CTAGAGTTTGGAGCAGGCGCTCAGATTGGCCAAAATAGGAAGGATGGATTCCACTCT 300

QY 301 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAGTCTGAGGCTTTTCCATCAGTAAT 360

Db 301 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAGTCTGAGGCTTTTCCATCAGTAAT 360

361 CAGTCCCTTTTGTCTTTCTTTTACGACCAACATGAACCTTGAGAACCCCTTAAGGTATA 420
Db 361 CAGTCCCTTTTGTCTTTCTTTTACGACCAACATGAACCTTGAGAACCCCTTAAGGTATA 420
QY 421 TCATTTAGTGAGTTGGCAGTTCCCAAGTGTCCCAAGAGGAGGCTGTTTGGTGGTGG 480
Db 421 TCATTTAGTGAGTTGGCAGTTCCCAAGTGTCCCAAGAGGAGGCTGTTTGGTGGTGG 480
QY 481 ATGGCCACTGTCAATTCCTGGTGTGATTTGTGAGAAAGTAAAGATACCCAGAAATGCCCTTCA 540
Db 481 ATGGCCACTGTCAATTCCTGGTGTGATTTGTGAGAAAGTAAAGATACCCAGAAATGCCCTTCA 540
QY 541 GGGAAACGTAAGCGTGTGTAACCAACCAAGAAACCTTCTTGGCCAACTGTGTGAC 600
Db 541 GGGAAACGTAAGCGTGTGTAACCAACCAAGAAACCTTCTTGGCCAACTGTGTGAC 600
QY 601 AAGGCGCTTTAAGTGTGAGAAATTAAGGTTTCACTCTCTACACAGAGAGGAGG 660
Db 601 AAGGCGCTTTAAGTGTGAGAAATTAAGGTTTCACTCTCTACACAGAGAGGAGG 660
QY 661 CCTTACAGTGCATACCAAGACTGCACCAAGGCGCTTGTCTTAAAGTACAAATACAA 720
Db 661 CCTTACAGTGCATACCAAGACTGCACCAAGGCGCTTGTCTTAAAGTACAAATACAA 720
QY 721 AGGCACATGGCTACTCTTCTCTGAGAAACCCCAAGTGTAAATTTTGTGAGAAATG 780
Db 721 AGGCACATGGCTACTCTTCTCTGAGAAACCCCAAGTGTAAATTTTGTGAGAAATG 780
QY 781 TTTTACCGGAAAGATCATCTGAAGATCACTCTTACACAGAGAGGAGGAGGAGG 840
Db 781 TTTTACCGGAAAGATCATCTGAAGATCACTCTTACACAGAGAGGAGGAGGAGG 840
QY 841 TTTTACCGGAAAGATCATCTGAAGATCACTCTTACACAGAGAGGAGGAGGAGG 900
Db 841 TTTTACCGGAAAGATCATCTGAAGATCACTCTTACACAGAGAGGAGGAGGAGG 900
QY 901 GCCTTCGATCGCCGCAACAAAGTGTGAGCTTCACTGTAAGAGTATGTTTGCAGAACTTTGAA 960
Db 901 GCCTTCGATCGCCGCAACAAAGTGTGAGCTTCACTGTAAGAGTATGTTTGCAGAACTTTGAA 960
QY 961 AGCAGCGGAGTGTCTTCTGAGAGCCTTAAATCTCATGCGGAGGAGGAGGAGGAGG 1020
Db 961 AGCAGCGGAGTGTCTTCTGAGAGCCTTAAATCTCATGCGGAGGAGGAGGAGGAGG 1020
QY 1021 AAGAGAAAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
Db 1021 AAGAGAAAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
QY 1081 CGGAGACACATGTTGTTGACACTGGAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
Db 1081 CGGAGACACATGTTGTTGACACTGGAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
QY 1141 AGATTTGGGCGAAGAGGATCACTGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
Db 1141 AGATTTGGGCGAAGAGGATCACTGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
QY 1201 CTGAAGGTCAAAACAGAAACAGTGTGATTTCTTGGACCCATTTACCTGCAATGTCTGTG 1260
Db 1201 CTGAAGGTCAAAACAGAAACAGTGTGATTTCTTGGACCCATTTACCTGCAATGTCTGTG 1260
QY 1261 CCTATAAAGACAGAGCTCTTCCGGTGTGATGTCCTTACCTTCCAGTGAACCTGTTATCAAG 1320
Db 1261 CCTATAAAGACAGAGCTCTTCCGGTGTGATGTCCTTACCTTCCAGTGAACCTGTTATCAAG 1320
QY 1321 CCATTCAAAACACTTTTGGAGTTAAACCTCTCAACACTCCATTTCAGTCCATCAGAGC 1380
Db 1321 CCATTCAAAACACTTTTGGAGTTAAACCTCTCAACACTCCATTTCAGTCCATCAGAGC 1380
QY 1381 TCGGGATCTGCCACCAAAATGATCAAACTTTTACCTTTGGGAGTATGATGCCAATAGAT 1440
Db 1381 TCGGGATCTGCCACCAAAATGATCAAACTTTTACCTTTGGGAGTATGATGCCAATAGAT 1440
QY 1441 ATGACACTGTGTTTCACTCTCACCACCTTTCTTTCAAAATATCCGTTCTACCTCA 1500

QY	61	GC	CGCGTCCAGCCCGGAAATATAGAAAAAAATATATAGAAATTC	CGCGGGCGGTGTAGAG	120
DB					
QY	61	GC	CGCGTCCAGCCCGGAAATATAGAAAAAAATATATAGAAATTC	CGCGGGCGGTGTAGAG	120
DB					
QY	121	GC	CGCGAGCGGGCGGAGGAGAGATGTTAAAGCCCGCGGTTCCTCTTGGTGTGCGCTT	180	
DB					
QY	121	GC	CGCGAGCGGGCGGAGGAGAGATGTTAAAGCCCGCGGTTCCTCTTGGTGTGCGCTT	180	
DB					
QY	181	GG	CGGTATTTGGCACCCAGAAATGCTTCACTCTGTGACGGTCTATTAATAAGGTTCGCTTG	240	
DB					
QY	181	GG	CGGTATTTGGCACCCAGAAATGCTTCACTCTGTGACGGTCTATTAATAAGGTTCGCTTG	240	
DB					
QY	241	CT	AGATTTGGAGCAGGCGCTCAGATGGCCAAAATGGGAAGATGGATTCACCTCTCT	300	
DB					
QY	241	CT	AGATTTGGAGCAGGCGCTCAGATGGCCAAAATGGGAAGATGGATTCACCTCTCT	300	
DB					
QY	301	TCC	ACGAAGATCAATGGGACTGGCTAAGATCAAGCTCTGAGGCTTTTCCATCAGTAAAT	360	
DB					
QY	301	TCC	ACGAAGATCAATGGGACTGGCTAAGATCAAGCTCTGAGGCTTTTCCATCAGTAAAT	360	
DB					
QY	361	CAG	TCCCTTTTGTCTTTTACGACCAATCAAACTTGAGAAGCCACCTAAAGCTATA	420	
DB					
QY	361	CAG	TCCCTTTTGTCTTTTACGACCAATCAAACTTGAGAAGCCACCTAAAGCTATA	420	
DB					
QY	421	TC	ATTAGTGGAGTTGGCAGTTCCTCCAAAGTGTCCAAAGAAGCGCTGGTTAGCTGG	480	
DB					
QY	421	TC	ATTAGTGGAGTTGGCAGTTCCTCCAAAGTGTCCAAAGAAGCGCTGGTTAGCTGG	480	
DB					
QY	481	AT	GGCCACTGTCTCTGTGTATTTGTCAAGTAGAGATACCCAGAAAGTCCCTTCA	540	
DB					
QY	481	AT	GGCCACTGTCTCTGTGTATTTGTCAAGTAGAGATACCCAGAAAGTCCCTTCA	540	
DB					
QY	541	GG	AAACGTAAAGCGTGGTGAACCAAAACCAAGAAAAAATCTTCTTGCCAACTGTGTAC	600	
DB					
QY	541	GG	AAACGTAAAGCGTGGTGAACCAAAACCAAGAAAAAATCTTCTTGCCAACTGTGTAC	600	
DB					
QY	601	AAG	GCCTTAAACAGTGTGGAAAAATTAAGGTTCACTCTCTCACAGAGAGAGG	660	
DB					
QY	601	AAG	GCCTTAAACAGTGTGGAAAAATTAAGGTTCACTCTCTCACAGAGAGAGG	660	
DB					
QY	661	CC	TACAAAGTGCATACAAAGACTGCACCAAGGCTTTGTTTCTAAGTACAAATACAA	720	
DB					
QY	661	CC	TACAAAGTGCATACAAAGACTGCACCAAGGCTTTGTTTCTAAGTACAAATACAA	720	
DB					
QY	721	AG	GCATGGCTACTCATTTCTTGAGAAAAACCAAGTGTAAATTTCTGAGAAAAATG	780	
DB					
QY	721	AG	GCATGGCTACTCATTTCTTGAGAAAAACCAAGTGTAAATTTCTGAGAAAAATG	780	
DB					
QY	781	TT	TCCCGGAAAGATCATCTGAAGATCACTCCATACACGACCCCTAACAAAGAGACG	840	
DB					
QY	781	TT	TCCCGGAAAGATCATCTGAAGATCACTCCATACACGACCCCTAACAAAGAGACG	840	
DB					
QY	841	TT	TAAAGTGGAGAAATGTGGCAAGAACTACAAATCAAGCTTGAGATTAACGCTCACTTG	900	
DB					
QY	841	TT	TAAAGTGGAGAAATGTGGCAAGAACTACAAATCAAGCTTGAGATTAACGCTCACTTG	900	
DB					
QY	901	GC	CTTCATGCGCGCAACAGTGTCACTCACTCTGAAGGTATGTTTGGCAAACTTTTGA	960	
DB					
QY	901	GC	CTTCATGCGCGCAACAGTGTCACTCACTCTGAAGGTATGTTTGGCAAACTTTTGA	960	
DB					
QY	961	AG	CAGGAGTGTCTCTGGAGCACCTTAAATCTCATGCAGCAAGTCTCTGTGGGGTT	1020	
DB					
QY	961	AG	CAGGAGTGTCTCTGGAGCACCTTAAATCTCATGCAGCAAGTCTCTGTGGGGTT	1020	
DB					
QY	1021	AA	AAAAAAGACCAAGTGGCAACATTTGTGATCGCCGGTTCTACACCCGAAAGGATGTC	1080	
DB					
QY	1021	AA	AAAAAAGACCAAGTGGCAACATTTGTGATCGCCGGTTCTACACCCGAAAGGATGTC	1080	
DB					
QY	1081	CG	GAGACATAGTGTGTGCACACTGGAAGAAAGGACTTCTCTGTCAAGTATTTGTGCACAG	1140	
DB					
QY	1081	CG	GAGACATAGTGTGTGCACACTGGAAGAAAGGACTTCTCTGTCAAGTATTTGTGCACAG	1140	
DB					
QY	1141	AG	ATTTGGCGAAAGGATCACCTTGACTCGACATATGAAGAAGAGTCACAATCAAGAGCTT	1200	

1141	AGATTGGCGAAGATCACTGACTGACATATGAAGAGAGTCACAATCAAGAGCTT	1200
1201	CTGAAGGTCAAAACAGAACCGAGTGATTCCTTGAGCCCAATTTACCTGCAATGTGTCGTG	1260
1201	CTGAAGGTCAAAACAGAACCGAGTGATTCCTTGAGCCCAATTTACCTGCAATGTGTCGTG	1260
1261	CCTATAAAGACGAGCTCCTCCGGTGATGTCCTTACCTTCAGTGAACTGTTACAAAG	1320
1261	CCTATAAAGACGAGCTCCTCCGGTGATGTCCTTACCTTCAGTGAACTGTTACAAAG	1320
1321	CCATTCAAAAACACTTTGCAAGTTAAACCTCTACACACTCCCAATTCAGTCCATGAGAGC	1380
1321	CCATTCAAAAACACTTTGCAAGTTAAACCTCTACACACTCCCAATTCAGTCCATGAGAGC	1380
1381	TCGGGATCTGCCCAACCAATGATCACAACTTTTACCTTTGGGAATGACATGCCCAATAGAT	1440
1381	TCGGGATCTGCCCAACCAATGATCACAACTTTTACCTTTGGGAATGACATGCCCAATAGAT	1440
1441	ATGGAACACTGTTCAATCCCTCTCACCACTTCTTTCAAAATATCCGTTACGTTCTACCTCA	1500
1441	ATGGAACACTGTTCAATCCCTCTCACCACTTCTTTCAAAATATCCGTTACGTTCTACCTCA	1500
1501	TATGCAATTTCTATTCCTGAAAAAGAACGCAATTAAGGGGGAATTCAGAGTTACCTG	1560
1501	TATGCAATTTCTATTCCTGAAAAAGAACGCAATTAAGGGGGAATTCAGAGTTACCTG	1560
1561	ATGGAGTTACAAGTGGCGTCCCTCTTCATCCCAAGATTCACAAGCATGTCATCATCT	1620
1561	ATGGAGTTACAAGTGGCGTCCCTCTTCATCCCAAGATTCACAAGCATGTCATCATCT	1620
1621	AAGCTAGGGTTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGAGACCTCTCCCTA	1680
1621	AAGCTAGGGTTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGAGACCTCTCCCTA	1680
1681	TCGAAAAGCTCTATCTCCATCAGTGACCCCTAAAACACACAGCAATTTGGAATTTCTCAG	1740
1681	TCGAAAAGCTCTATCTCCATCAGTGACCCCTAAAACACACAGCAATTTGGAATTTCTCAG	1740
1741	TTGTTTAATTTCAATACCTTTAAATGGTCTCCCTCTATAATCCTCTATCAGTGGGAGCCTT	1800
1741	TTGTTTAATTTCAATACCTTTAAATGGTCTCCCTCTATAATCCTCTATCAGTGGGAGCCTT	1800
1801	GGAAATGAGCTAATCCCAGGAAGAACATTCCTTCGTCTCCAGCTCCCCACACAAACA	1860
1801	GGAAATGAGCTAATCCCAGGAAGAACATTCCTTCGTCTCCAGCTCCCCACACAAACA	1860
1861	CAGGATCTTCAGGATCTCGAAACACATATAGGGCTTGGGTCTCTGCACTCACTGTGAGCA	1920
1861	CAGGATCTTCAGGATCTCGAAACACATATAGGGCTTGGGTCTCTGCACTCACTGTGAGCA	1920
1921	GCTTTCCACGACGTTTAAGCACAAAGTACCACCTCCCACTGTTCCCATCAAGCTTTTCAG	1980
1921	GCTTTCCACGACGTTTAAGCACAAAGTACCACCTCCCACTGTTCCCATCAAGCTTTTCAG	1980
1981	TAGGATTCCTGGGACATGGATTCAATACAGAAATGATGTAGCTGGCCCTAGATGACC	2040
1981	TAGGATTCCTGGGACATGGATTCAATACAGAAATGATGTAGCTGGCCCTAGATGACC	2040
2041	ATTTTTATTTTAGTGCCCTACTTTAAAAACAGTATAAAAAATTTCTGCTTTTGTATAATCAA	2100
2041	ATTTTTATTTTAGTGCCCTACTTTAAAAACAGTATAAAAAATTTCTGCTTTTGTATAATCAA	2100
2101	ATTTTTCATTAAAGCCAGTATAAAATAGAAACTAGCTTTTAACTGAGCTTTGGAAACCATTT	2160
2101	ATTTTTCATTAAAGCCAGTATAAAATAGAAACTAGCTTTTAACTGAGCTTTGGAAACCATTT	2160
2161	GTGTTTCAGTTAAGTTTACCTGGGTATTTTGTCCCTGATTCATGCGCAATTTGTCAATTTTA	2220
2161	GTGTTTCAGTTAAGTTTACCTGGGTATTTTGTCCCTGATTCATGCGCAATTTGTCAATTTTA	2220
2221	AGACTTTTTTTTTTCCATATAGGAAAGCCATTAATAGTAGTAAACTTTTACAAATCCCA	2280

Db 2221 AGACCTTTTTCATATAGGAAAGCCATTATTAGTAGTAAACCTTTTACAAATCCCA 2280
Qy 2281 TTTTCAAAATTAATTTTAGATCTTAAAAATTTTCAATTTTGTCTAAATAACAGTGGCTTACC 2340
Db 2281 TTTTCAAAATTAATTTTAGATCTTAAAAATTTTCAATTTTGTCTAAATAACAGTGGCTTACC 2340
Qy 2341 TTTTGACATCTGGCTCAATTAATAATTTTAGCAATAGAAATGTAATTTGATAAAAAGTTTG 2400
Db 2341 TTTTGACATCTGGCTCAATTAATAATTTTAGCAATAGAAATGTAATTTGATAAAAAGTTTG 2400
Qy 2401 TGAATAACTCAAGGTTTAAATTTTCTTACTAGCTTCTAAATGCAATTAATCAAGTGC 2460
Db 2401 TGAATACTCAAGGTTTAAATTTTCTTACTAGCTTCTAAATGCAATTAATCAAGTGC 2460
Qy 2461 TTCAAAATGAATTAAGAGTCCAGTTTCGGAAGATATAAATGTTTGTAGATACACCAATAA 2520
Db 2461 TTCAAAATGAATTAAGAGTCCAGTTTCGGAAGATATAAATGTTTGTAGATACACCAATAA 2520
Qy 2521 TTTTCAGATCAGTATATTCTGAAGACTCTCTGTTGTCTGGCTAAATATTGTCATCTTTA 2580
Db 2521 TTTTCAGATCAGTATATTCTGAAGACTCTCTGTTGTCTGGCTAAATATTGTCATCTTTA 2580
Qy 2581 TTATGAGCCCTTAAAGGAACCAACCCCTAAACCAAGCAATCAATTTATAGCAAAAG 2640
Db 2581 TTATGAGCCCTTAAAGGAACCAACCCCTAAACCAAGCAATCAATTTATAGCAAAAG 2640
Qy 2641 AGACTCTGTTAGTGACATGGCAATTTCTGTCATCTTAATAGTTGGCCCTAAATTTAGTACA 2700
Db 2641 AGACTCTGTTAGTGACATGGCAATTTCTGTCATCTTAATAGTTGGCCCTAAATTTAGTACA 2700
Qy 2701 CAGGATATTTTGTGCTGTTTCATCTCTTAAACATGCTATCTTTTCAATTAATAAGTA 2760
Db 2701 CAGGATATTTTGTGCTGTTTCATCTCTTAAACATGCTATCTTTTCAATTAATAAGTA 2760
Qy 2761 ATAGTGTATGGCAATTCGGGCTTTCAGAGTGCATATATAGTAGATCTCTTTAGTCTTTTC 2820
Db 2761 ATAGTGTATGGCAATTCGGGCTTTCAGAGTGCATATATAGTAGATCTCTTTAGTCTTTTC 2820
Qy 2821 CACCTTTTCATCCAAAGGGTGGTCAAGTGCAGCCAGCAATTTATTTTCAATTTGGGC 2880
Db 2821 CACCTTTTCATCCAAAGGGTGGTCAAGTGCAGCCAGCAATTTATTTTCAATTTGGGC 2880
Qy 2881 CACGGTTAGTCCATATCTAGAGCCATTTGTGGAATCTGCAGCCATGAGTGTGTTATCCC 2940
Db 2881 CACGGTTAGTCCATATCTAGAGCCATTTGTGGAATCTGCAGCCATGAGTGTGTTATCCC 2940
Qy 2941 ACAGTGGATTGACTCAGCCCTCTGCGGTGACAGACTTCTAAGCAGGAATAGACGTGAA 3000
Db 2941 ACAGTGGATTGACTCAGCCCTCTGCGGTGACAGACTTCTAAGCAGGAATAGACGTGAA 3000
Qy 3001 GCACATGGTTACATTTGGGAATTTGTGTAGGATCATGGCCCTGTAGCCAGGTTAAAA 3060
Db 3001 GCACATGGTTACATTTGGGAATTTGTGTAGGATCATGGCCCTGTAGCCAGGTTAAAA 3060
Qy 3061 ACTGCACATTTTAGAAGTAAAGTAAAGCATAGCGCTTATATCAATTTCTGCTGAATTTG 3120
Db 3061 ACTGCACATTTTAGAAGTAAAGTAAAGCATAGCGCTTATATCAATTTCTGCTGAATTTG 3120
Qy 3121 ATATGTTTTTCTTCCCTTAAAGTAAAGTAAAGCATAGCGCTTATATCAATTTCTGCTGAATTTG 3180
Db 3121 ATATGTTTTTCTTCCCTTAAAGTAAAGTAAAGCATAGCGCTTATATCAATTTCTGCTGAATTTG 3180
Qy 3181 ATGTTATCTTCTGATTCGAATCCATCTTTCTTGGCAATTTTGGATGAGAT 3240
Db 3181 ATGTTATCTTCTGATTCGAATCCATCTTTCTTGGCAATTTTGGATGAGAT 3240
Qy 3241 TTGAAGTTTAAATGCAATTAAGAACTACCTGATGAATACCAAAAGTTTAAAGTACCTAG 3300
Db 3241 TTGAAGTTTAAATGCAATTAAGAACTACCTGATGAATACCAAAAGTTTAAAGTACCTAG 3300
Qy 3301 AAATATATACAGTAAATCCCACTTTTCATGCTCTCTGGGAAATGATAGGATATTGCAA 3360
Db 3301 AAATATATACAGTAAATCCCACTTTTCATGCTCTCTGGGAAATGATAGGATATTGCAA 3360

Qy 3361 ATAAAGTTGAGTTTGTAGAGGGTAAACAAGTAAAGTAAACAAACCTATCTTTGGTTAAACAT 3420
Db 3361 ATAAAGTTGAGTTTGTAGAGGGTAAACAAGTAAAGTAAACAAACCTATCTTTGGTTAAACAT 3420
Qy 3421 GAAATAAACAAATTCAGAAATATATTATATCTCACTCAATAATTTATAGGCTTTTCTTCAAT 3480
Db 3421 GAAATAAACAAATTCAGAAATATATTATATCTCACTCAATAATTTATAGGCTTTTCTTCAAT 3480
Qy 3481 AGACAACCAACATATCTTTAAAGGCTTAAATTAATATATTTTCTAAAGGCTCAGTTGG 3540
Db 3481 AGACAACCAACATATCTTTAAAGGCTTAAATTAATATATTTTCTAAAGGCTCAGTTGG 3540
Qy 3541 GACATTAACCTTAAGAAACATATCTTATTAAGCACTTTTAAAGGCTTATCTTAAACCACT 3600
Db 3541 GACATTAACCTTAAGAAACATATCTTATTAAGCACTTTTAAAGGCTTATCTTAAACCACT 3600
Qy 3601 TCCGTTGGGATCGGGCAAGGTTGGGAGGTTTGAAGAGTATATATCTTCTTAAAAA 3660
Db 3601 TCCGTTGGGATCGGGCAAGGTTGGGAGGTTTGAAGAGTATATATCTTCTTAAAAA 3660
Qy 3661 AAAACAGAAAGAAAAATATTTCTGAGCACTCATTTAGCCCTATATGGAACCTTCTTCCCT 3720
Db 3661 AAAACAGAAAGAAAAATATTTCTGAGCACTCATTTAGCCCTATATGGAACCTTCTTCCCT 3720
Qy 3721 TTTGTAGGCCAGTTTATCTGAGCACTTGAATTTTACCAAGAAATTTCTAAAAATGAGT 3780
Db 3721 TTTGTAGGCCAGTTTATCTGAGCACTTGAATTTTACCAAGAAATTTCTAAAAATGAGT 3780
Qy 3781 GCAGATTACTGAATATAATACATTTTAAATATTTGGGAGTAGTATATTTTGTGAGA 3840
Db 3781 GCAGATTACTGAATATAATACATTTTAAATATTTGGGAGTAGTATATTTTGTGAGA 3840
Qy 3841 AATGTAATTTGTAATTAATGTAATGGGGCTTCAATATATATATATATATATACACACAC 3900
Db 3841 AATGTAATTTGTAATTAATGTAATGGGGCTTCAATATATATATATATATACACACAC 3900
Qy 3901 ACACATGACACACATACCGCACTTCAATAGAAATCAAAGTTGCTCTGGAAGGAGCTTTGG 3960
Db 3901 ACACATGACACACATACCGCACTTCAATAGAAATCAAAGTTGCTCTGGAAGGAGCTTTGG 3960
Qy 3961 CTCCTGATATTTTATCATGCTCTATTTTAAATCTTGGAGCAGTAGTTTATATAC 4020
Db 3961 CTCCTGATATTTTATCATGCTCTATTTTAAATCTTGGAGCAGTAGTTTATATAC 4020
Qy 4021 TTATGATTTTAAATTTTATTAATGAAATTAATTTTAAATGAGTGTGTTCCAAAGGC 4080
Db 4021 TTATGATTTTAAATTTTATTAATGAAATTAATTTTAAATGAGTGTGTTCCAAAGGC 4080
Qy 4081 ATTAATTTATATATTTTAAATGAAAGTACATTTTAAATTTTCAAACTGCTCTAGC 4140
Db 4081 ATTAATTTATATATTTTAAATGAAAGTACATTTTAAATTTTCAAACTGCTCTAGC 4140
Qy 4141 TTTTGAATGAGGAGATATTTTCTGAAAGTAGGCTTTTGGCTCTGCTTCAATCTGCTT 4200
Db 4141 TTTTGAATGAGGAGATATTTTCTGAAAGTAGGCTTTTGGCTCTGCTTCAATCTGCTT 4200
Qy 4201 CTTTGTAGTTTCTTGAAGACAGATTTTCTTACCTTAAATCTTTAGTTGAATGATGTTGTTCA 4260
Db 4201 CTTTGTAGTTTCTTGAAGACAGATTTTCTTACCTTAAATCTTTAGTTGAATGATGTTGTTCA 4260
Qy 4261 ATATGCTTTTAAATCCATATAAAGGAAATTTGGTGAAGAGTAAATGAGGAGTAAAC 4320
Db 4261 ATATGCTTTTAAATCCATATAAAGGAAATTTGGTGAAGAGTAAATGAGGAGTAAAC 4320
Qy 4321 CTATTTTAAATAGAAATCAAAATAGCAAGTGTGGAAGCACTTCTTATCTGTTTAAA 4380
Db 4321 CTATTTTAAATAGAAATCAAAATAGCAAGTGTGGAAGCACTTCTTATCTGTTTAAA 4380
Qy 4381 ATGACTTTAAGAGTCAATCAATTTAGTGAATGAGCACTTCTGAGCACTTCTAGGCTGATTC 4440
Db 4381 ATGACTTTAAGAGTCAATCAATTTAGTGAATGAGCACTTCTGAGCACTTCTAGGCTGATTC 4440


```
Db 6601 CATGTGCACTTGC AAAAGGAGATCTCGGACATGCAAAATGTTACCGAGACAAACCCAGCTT 6660
QY 6661 TTGTCCACAAGGTGACTGTAACTCAAGATGGAAGTGGGCTTTATATAGGGTGTGGACT 6720
Db 6661 TTGTCCACAAGGTGACTGTAACTCAAGATGGAAGTGGGCTTTATATAGGGTGTGGAGT 6720
QY 6721 GAAGAACATGCTGTATGTTACTAAACAGCCCTTTGAATTTTAAACAAAACCTGGGAATCCATT 6780
Db 6721 GAAGAACATGCTGTATGTTACTAAACAGCCCTTTGAATTTTAAACAAAACCTGGGAATCCATT 6780
QY 6781 AGGAAACGATGTCATCATCTACCTGAACATAGCTGACGTGCTGAAATTTGTATTTTGTAGCT 6840
Db 6781 AGGAAACGATGTCATCATCTACCTGAACATAGCTGACGTGCTGAAATTTGTATTTTGTAGCT 6840
QY 6841 AATGAAAAAGTGTGTCGACTAGTACTCTAAATAATGTTCTAAATGATAAAGTGTGAGTCAA 6900
Db 6841 AATGAAAAAGTGTGTCGACTAGTACTCTAAATAATGTTCTAAATGATAAAGTGTGAGTCAA 6900
QY 6901 AATGAAAAAGAAAAAATCTGCATTCAGCCCGCAATTTTGTATATTTTATTTGCAATTTAA 6960
Db 6901 AATGAAAAAGAAAAAATCTGCATTCAGCCCGCAATTTTGTATATTTTATTTGCAATTTAA 6960
QY 6961 AATTCCTATTCTGTAATATTTGGGAATCAAGTGGCTTATCATGTATATCGTGTACTTAA 7020
Db 6961 AATTCCTATTCTGTAATATTTGGGAATCAAGTGGCTTATCATGTATATCGTGTACTTAA 7020
QY 7021 ATGTATTACAAACTACTGTTGTTATTTGTAATAATATAGACAAAGATCATATTTTGT 7080
Db 7021 ATGTATTACAAACTACTGTTGTTATTTGTAATAATATAGACAAAGATCATATTTTGT 7080
QY 7081 GTGTGTATAAGCTCTGTAATAATAGCAATACATTTATGAAGCTGCAGTATACATATTT 7140
Db 7081 GTGTGTATAAGCTCTGTAATAATAGCAATACATTTATGAAGCTGCAGTATACATATTT 7140
QY 7141 AAACATTCACATCCAAAGACAGACTATTTATGTCGCATATACAGATTTAAATATTA 7200
Db 7141 AAACATTCACATCCAAAGACAGACTATTTATGTCGCATATACAGATTTAAATATTA 7200
QY 7201 ATTTCGCTGTAATAATAATAGTACTGAGCTTCTGTGGCTACAGTGTATGTTGCG 7260
Db 7201 ATTTCGCTGTAATAATAATAGTACTGAGCTTCTGTGGCTACAGTGTATGTTGCG 7260
QY 7261 TGTAAGAATAAGATATGTAATTCACAAAATATATGTAATAAATCTCGTGC 7313
Db 7261 TGTAAGAATAAGATATGTAATTCACAAAATATATGTAATAAATCTCGTGC 7313

RESULT 3
AC107952/c
LOCUS AC107952 Homo sapiens chromosome 8, clone RP11-140I16, complete sequence.
ACCESSION AC107952
VERSION AC107952.5 GI:21930261
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 142102)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-140I16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 142102)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,l,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
```

```
TITLE JOURNAL
REFERENCE AUTHORS
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 142102)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,l,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagob,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 142102)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,l,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagob,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 23, 2002 this sequence version replaced gi:21699524.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24490
Center clone name: 140_I_16
----- Location/Qualifiers
```

FEATURES

Db 30559 TGCCTATAAAGACGAGCTCCTTCGGTGATGTCCTTACCTCCAGTGAACCTGTTATCAA 30500
Qy 1319 AGCCATTCAAAACACTTTGCACTTAAACCTCTACAAACACTCCATTTAGTCCATGCAGA 1378
Db 30499 AGCCATTCAAAACACTTTGCACTTAAACCTCTACAAACACTCCATTTAGTCCATGCAGA 30440
Qy 1379 GCTGGGATCTGCCCAACAAATGATCACAACTTTTACCTTTGGGAATGACATGCCCAATAG 1438
Db 30439 GCTGGGATCTGCCCAACAAATGATCACAACTTTTACCTTTGGGAATGACATGCCCAATAG 30380
Qy 1439 ATATGGACACTGTTGATCCCTCTCACCACCTTTCTTTTCAAAATGATCGTTTACCT 1498
Db 30379 ATATGGACACTGTTGATCCCTCTCACCACCTTTCTTTTCAAAATGATCGTTTACCT 30320
Qy 1499 CATATGCAATTTCTATTTCTGAAAAGACAGCCATTAAAGGGGGAATTTGAGAGTTACC 1558
Db 30319 CATATGCAATTTCTATTTCTGAAAAGACAGCCATTAAAGGGGGAATTTGAGAGTTACC 30260
Qy 1559 TGATGGAGTTACAAGTGGCGTGCCTCTTTCATCCCAAGATTCTCAAGCATCGTCATCAT 1618
Db 30259 TGATGGAGTTACAAGTGGCGTGCCTCTTTCATCCCAAGATTCTCAAGCATCGTCATCAT 30200
Qy 1619 CTAAGCTAGGTTGATCCTCAGATTGGTCCCTAGATGATGGTGAGAGACCTCTCC 1678
Db 30199 CTAAGCTAGGTTGATCCTCAGATTGGTCCCTAGATGATGGTGAGAGACCTCTCC 30140
Qy 1679 TATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACACAGCATTTGATTTTCTC 1738
Db 30139 TATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACAGCATTTGATTTTCTC 30080
Qy 1739 AGTTGTTTAAATTCATACCTTTTAAATGGTCTCCCTATAATCTCTATCAGTGGGAGCC 1798
Db 30079 AGTTGTTTAAATTCATACCTTTTAAATGGTCTCCCTATAATCTCTATCAGTGGGAGCC 30020
Qy 1799 TTGGAATGAGCTATTTCCAGGAGAGACACATCTTCTGTTTCCAGCTGCCCAACAAA 1858
Db 30019 TTGGAATGAGCTATTTCCAGGAGAGAGACACATCTTCTGTTTCCAGCTGCCCCACAAA 29960
Qy 1859 CACAGGATCTTCAGGATCCTGCAACACTATAGGGCTTTGGGTCTCTGCACCTACCTGTCAG 1918
Db 29959 CACAGGATCTTCAGGATCCTGCAACACTATAGGGCTTTGGGTCTCTGCACCTACCTGTCAG 29900
Qy 1919 CAGCTTCCAGCAGAGTTTAAAGCACAGTACACCCCTCCACGTTTCCATCAAGCTTTTC 1978
Db 29899 CAGCTTCCAGCAGAGTTTAAAGCACAGTACACCCCTCCACGTTTCCATCAAGCTTTTC 29840
Qy 1979 AGTAGGATCTGGGACATGGAATCATTACAGAAATGATGTAGCTGTGCCCTAGATGA 2038
Db 29839 AGTAGGATCTGGGACATGGAATCATTACAGAAATGATGTAGCTGTGCCCTAGATGA 29780
Qy 2039 CCATTTTATTTTATGTCCTACTTTTAAACAGTATAAAATTTCTGCTTTTGTATATATAC 2098
Db 29779 CCATTTTATTTTATGTCCTACTTTTAAACAGTATAAAATTTCTGCTTTTGTATATATAC 29720
Qy 2099 AAATTTTCATTAAGCCAGTATAAATAGAACTAGCTTTTAAACTGAGCTTTGGAACCAT 2158
Db 29719 AAATTTTCATTAAGCCAGTATAAATAGAACTAGCTTTTAAACTGAGCTTTGGAACCAT 29660
Qy 2159 TTGTTTCAGTTAAGTTTACCTGGGTATTTTGTCTGATTCACCTGCCAATTTGTACATTT 2218
Db 29659 TTGTTTCAGTTAAGTTTACCTGGGTATTTTGTCTGATTCACCTGCCAATTTGTACATTT 29600
Qy 2219 TAAGACTTTTTTTTTTCCATATAGGAAGCCATTATTAGTAGTAACCTTTTACAAATCC 2278
Db 29599 TAGAC - TTTT TTTT TTTTCCATATAGGAAGCCATTATTAGTAGTAACCTTTTACAAATCC 29541
Qy 2279 CATTTTCAAAATTTACTTTTATGATCTTAAATTTTCAATTTTCTGCTTAAATAGAGTGCCTCTA 2338
Db 29540 CATTTTCAAAATTTACTTTTATGATCTTAAATTTTCAATTTTCTGCTTAAATAGAGTGCCTCTA 29481
Qy 2339 CCTTTTGACATCTGGCTCATTTAAATAATTTAGCAATAGAAATGTAATTTGTAATAAAAGTT 2398
Db 29480 CCTTTTGACATCTGGCTCATTTAAATAATTTAGCAATAGAAATGTAATTTGTAATAAAAGTT 29421

Qy 2399 TGTGAATTAACCTCAAGGTTTAAATTTTCTTACTAGCTTCTAAATGGAATTAATAATCAAGT 2458
Db 29420 TGTGAATTAACCTCAAGGTTTAAATTTTCTTACTAGCTTCTAAATGGAATTAATAATCAAGT 29361
Qy 2459 GCTTCAAAATGAATTAAGAGTCCAGTTTCGGAAGATAAATAATGTTTGTAGTACACCAT 2518
Db 29360 GCTTCAAAATGAATTAAGAGTCCAGTTTCGGAAGATAAATAATGTTTGTAGTACACCAT 29301
Qy 2519 AATTTTCAGATCAGTATATTTCTGAAGACTCTCTGTGTCTGTGGCTAAATAATTTGCCATCTT 2578
Db 29300 AATTTTCAGATCAGTATATTTCTGAAGACTCCCTGTGTCTGTGGCTAAATAATTTGCCATCTT 29241
Qy 2579 TATTTATGAGCCCTTTTAAAGGAAAACAAACCCCTAAACACAAAGCATCAGTATTTATAGCAAA 2638
Db 29240 TATTTATGAGCCCTTTTAAAGGAAAACAAACCCCTAAACACAAAGCATCAGTATTTATAGCAAA 29181
Qy 2639 AGAGACTCTGTTAGTGACATGGCATTTCTGTCTCACTTAATAGTTGGCCCTAAATTTAGTA 2698
Db 29180 AGAGACTCTGTTAGTGACATGGCATTTCTGTCTCACTTAATAGTTGGCCCTAAATTTAGTA 29121
Qy 2699 CACAGGATATTTTGTGTTTTCATCTTTTAAACATGCTATCTTTTCAATTTAAATAATAG 2758
Db 29120 CACAGGATATTTTGTGTTTTCATCTTTTAAACATGCTATCTTTTCAATTTAAATAATAG 29061
Qy 2759 TATTTAGTATGGCATTTGGGTCCTTCAGAGTCGATATATAGTAGATCTCTTTAGTCTTT 2818
Db 29060 TATTTAGTATGGCATTTGGGTCCTTCAGAGTCGATATATAGTAGATCTCTTTAGTCTTT 29001
Qy 2819 TCCACCTTTTCATCCCAAGGGTGGTCAAGTGCAGCAGCAATTTATTTTCAATTTGTTGG 2878
Db 29000 TCCACCTTTTCATCCCAAGGGTGGTCAAGTGCAGCAGCAATTTATTTTCAATTTGTTGG 28941
Qy 2879 CCCACGGTTAGTCCATAATCTTAGAGCCATTGTGGAACCTGCAGCCATGAGGTGTGTTATC 2938
Db 28940 CCCACGGTTAGTCCATAATCTTAGAGCCATTGTGGAACCTGCAGCCATGAGGTGTGTTATC 28881
Qy 2939 CCACAGTGATTTGACTCAGCCTCTGTGGGTGACAGACTTTTAAAGCAGGAAGATAGACGTG 2998
Db 28880 CCACAGTGATTTGACTCAGCCTCTGTGGGTGACAGACTTTTAAAGCAGGAAGATAGACGTG 28821
Qy 2999 AAGCACATGTTTACATTTTGGGAACTTGTGTAGGATCATGGCCCTGTAGCCAGGTTAA 3058
Db 28820 AAGCACATGTTTACATTTTGGGAACTTGTGTAGGATCATGGCCCTGTAGCCAGGTTAA 28761
Qy 3059 AAACCTGGACTTTTATAGAGTAAAGTAAAGCATATCGCTTATATCATTTCTTGTGAAT 3118
Db 28760 AAACCTGGACTTTTATAGAGTAAAGTAAAGCATATCGCTTATATCATTTCTTGTGAAT 28701
Qy 3119 TGATATGTTTTTCTTCCCTTAAAGATCAAAAGCAGAAAAACAAACACAGTCTACTC 3178
Db 28700 TGATATGTTTTTCTTCCCTTAAAGATCAAAAGCAGAAAAACAAACACAGTCTACTC 28641
Qy 3179 CGATGTTATCTTCTGATTCATCTGATCCATCTTCTTCTGCAATATTTTGGATGGAGA 3238
Db 28640 CGATGTTATCTTCTGATTCATCTGATCCATCTTCTTCTGCAATATTTTGGATGGAGA 28581
Qy 3239 ATTTGGAAGTTAAATGCAATTTAGAAAACCTCTGATGAACCTACCAACAAAGTTTAAAGTACT 3298
Db 28580 ATTTGGAAGTTAAATGCAATTTAGAAAACCTCTGATGAACCTACCAACAAAGTTTAAAGTACT 28521
Qy 3299 AGAATATATACAGTAAATCCCTTTTCATGCACTCTCGGAAATGATAGAGTATTGC 3358
Db 28520 AGAATATATACAGTAAATCCCTTTTCATGCACTCTCGGAAATGATAGAGTATTGC 28461
Qy 3359 AAATAAGTTCAGTTTGTAGAGGGTAAACAAAGTAAAGTAAACCACTATCTTGGTTAAC 3418
Db 28460 AAATAAGTTCAGTTTGTAGAGGGTAAACAAAGTAAAGTAAACCACTATCTTGGTTAAC 28401
Qy 3419 ATGAAAATAACAATTTGAGAATATATATATTTCTGTAATTAATTAAGCTTTTCTCCTACA 3478
Db 28400 ATGAAAATAACAATTTGAGAATATATATATTTCTGTAATTAATTAAGCTTTTCTCCTACA 28341

QY	3479	TTAGACAACCAACATACTCTTAAAGGTCTAAATTAATATATATTTTCTAAGGGTCAGTT	3538	Db	27262	ACACACACACATACATTAATCTTGCCCTGCATGAAATTCAAATACATGAGGCACATCTT	27203
Db	28340	TTAGACAACCAACATACTCTTAAAGGTCTAAATTAATATATATTTTCTAAGGGTCAGTT	28281	QY	4619	CAGGGCACCAGTGTAAAAATTTTGAGTCTTAATTTTCATGTGTACACCTCTTTGCTCTG	4678
QY	3539	GGGACATTAACCTAAGAAACATATCTATTAAAGCATTGTTAAACACCTTATTTTAGGACCC	3598	Db	27202	CAGGGCACCAGTGTAAAAATTTTGAGTCTTAATTTTCATGTGTACACCTCTTTGCTCTG	27143
Db	28280	GGGACATTAACCTAAGAAACATATCTATTAAAGCATTGTTAAACACCTTATTTTAGGACCC	28221	QY	4679	TCCCAACCCACAGCTTGAATAACACTTCAGAGTAAGAGGGAATTCAGCTAATTTGTTTTT	4738
QY	3599	TTTCCGTTGGGATGGGGCAAGGGTGGAGGTTTTTGAAGATATATATCTCTTTAAA	3658	Db	27142	TCCCAACCCACAGCTTGAATAACACTTCAGAGTAAGAGGGAATTCAGCTAATTTGTTTTT	27083
Db	28220	TTTCCGTTGGGATGGGGCAAGGGTGGAGGTTTTTGAAGATATATATCTCTTTAAA	28161	QY	4739	TAAAATTCAGCTAGTGGTCACTAAACCCCTTTTGAGGAATTTCTATTAAAGATGAGGC	4798
QY	3659	AAAAACAGAAAGAAAATTTCTGAGCACTCATTTAGCCCTATATGGAACCTCTTTCC	3718	Db	27082	TAAAATTCAGCTAGTGGTCACTAAACCCCTTTTGAGGAATTTCTATTAAAGATGAGGC	27023
Db	28160	AAAAACAGAAAGAAAATTTCTGAGCACTCATTTAGCCCTATATGGAACCTCTTTCC	28101	QY	4799	AGACTCGCTTATTTGAATTCGCAATGTTCTAACAGGATGTAAACACAGAAATGGCTTTT	4858
QY	3719	TTTTTGTAGGGCAGTTTCACTGCAAGTTGCAATGTTTACCAGAAATTTCTAAATAATGA	3778	Db	27022	AGACTCGCTTATTTGAATTCGCAATGTTCTAACAGGATGTAAACACAGAAATGGCTTTT	26963
Db	28100	TTTTTGTAGGGCAGTTTCACTGCAAGTTGCAATGTTTACCAGAAATTTCTAAATAATGA	28041	QY	4859	TTTTCCCTAGAAAAAGATTTGTTTTCTATGTCAACTAGATATGATTTAAAAATAAGTAT	4918
QY	3779	GTGCAGATTACTGAATATATACATTTTAAATATTTGGAGTAGTATATTTGTTGA	3838	Db	26962	TTTTCCCTAGAAAAAGATTTGTTTTCTATGTCAACTAGATATGATTTAAAAATAAGTAT	26903
Db	28040	GTGCAGATTACTGAATATATACATTTTAAATATTTGGAGTAGTATATTTGTTGA	27981	QY	4919	TGCCAATGCTGTTTTCAATCTCTAGTGGCCAGAACTATTCCTTGAAATTTCTGTGTAGT	4978
QY	3839	GAAATGTAATTTGAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG	3898	Db	26902	TGCCAATGCTGTTTTCAATCTCTAGTGGCCAGAACTATTCCTTGAAATTTCTGTGTAGT	26843
Db	27980	GAAATGTAATTTGAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG	27922	QY	4979	GCCTTAGCTTGGTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	5038
QY	3899	ACACACATGACACATACCGCACTTCATAGAAATCAAAAGTTGCTCTCTGAAGGAGCTTT	3958	Db	26842	GCCTTAGCTTGGTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	26784
Db	27921	ACACACATGACACATACCGCACTTCATAGAAATCAAAAGTTGCTCTCTGAAGGAGCTTT	27862	QY	5039	AGTAGTTTGAAGATTTGGGCTCAGACAGATATGAACTCAATTCAGTTTCACTCCACA	5098
QY	3959	GGCTCCTGATATTTTATCATGCTCTATATTTTAACTTGGAGCAGTAGTTTTAT	4018	Db	26783	AGTAGTTTGAAGATTTGGGCTCAGACAGATATGAACTCAATTCAGTTTCACTCCACA	26724
Db	27861	GGCTCCTGATATTTTATCATGCTCTATATTTTAACTTGGAGCAGTAGTTTTAT	27802	QY	5099	TGTATGTACAAGTTAGTTCACCAACACCGAAGT-TGAGTGTGGAAGAGTCTTTGGCACTG	5157
QY	4019	ACTTATGTTTAAATTTTATATGAAAAATACATTTTAAAAAAGTGTGTTCCAAAG	4078	Db	26723	TGTATGTACAAGTTAGTTCACCAACACCGAAGTGTGAGTGTGGAAGAGTCTTTGGCACTG	26664
Db	27801	ACTTATGTTTAAATTTTATGAAAAATACATTTTAAAAAAGTGTGTTCCAAAG	27742	QY	5158	TAAAGCAATGCTATCCATTCATGTATACAGTACCTTTATAGTTATCGATCACTGTTAAAA	5217
QY	4079	GCATTAATAATPATATGTTAATPAAGGAAGTACATTTTAAATTTTCAAACTGCTCCTA	4138	Db	26663	TAAAGCAATGCTATCCATTCATGTATACAGTACCTTTATAGTTATCGATCACTGTTAAAA	26604
Db	27741	GCATTAATAATPATATGTTAATPAAGGAAGTACATTTTAAATTTTCAAACTGCTCCTA	27682	QY	5218	CTTTCAATTTTAAATCTTATACCAAGTTTCAAGTTTAAAAAATTCAAATGTCTGGCTG	5277
QY	4139	GCCTTTGATAGGAGATATTTTCTGAAAGTAGGCTTTGCTCTGCTCTCAATTAATGCTC	4198	Db	26603	CTTTCAATTTTAAATCTTATACCAAGTTTCAAGTTTAAAAAATTCAAATGTCTGGCTG	26544
Db	27681	GCCTTTGATAGGAGATATTTTCTGAAAGTAGGCTTTTCTGCTCTCAATTAATGCTC	27622	QY	5278	ATPATGCATCACTCTCTGTGCAACTTTTTTATTCATTTTAGTGTGTTCTTTCAAGCTGTGT	5337
QY	4199	TTCTTTTAGTTTCTATGAAACAGATTGCTTACCTAAATCTTTAGTTGAATGATTAGTGT	4258	Db	26543	ATPATGCATCACTCTCTGTGCAACTTTTTTATTCATTTTAGTGTGTTCTTTCAAGCTGTGT	26484
Db	27621	TTCTTTTAGTTTCTATGAAACAGATTGCTTACCTAAATCTTTAGTTGAATGATTAGTGT	27562	QY	5338	ATTTTGGCTATTTGTTGCTTGTGCTTTATTTTCTTAGTCATTTCTGGAAATATAGTAT	5397
QY	4259	CAATATTTGCTTTAATCACCATATAAAGGAAAAAATTTGGTGTACAGAGCAAAATAGAAA	4318	Db	26483	ATTTTGGCTATTTGTTGCTTGTGCTTTATTTTCTTAGTCATTTCTGGAAATATAGTAT	26424
Db	27561	CAATATTTGCTTTAATCACCATATAAAGGAAAAAATTTGGTGTACAGAGCAAAATAGAAA	27502	QY	5398	ATAATGTGTAAATTTGGACAGTAGCGGTTTTTAAAAACCATATACCTGACTGAAACATGAG	5457
QY	4319	ACCTATTTTAAATGAAATCAAAATAGCAAGTGTGGAGCACTATTATTTCTCTTTA	4378	Db	26423	ATAATGTGTAAATTTGGACAGTAGCGGTTTTTAAAAACCATATACCTGACTGAAACATGAG	26364
Db	27501	ACCTATTTTAAATGAAATCAAAATAGCAAGTGTGGAGCACTATTATTTCTCTTTA	27442	QY	5458	CCAGAGCCGATTGCTTTTAAAGCTAATAATGAATGTTTAAAGAGTACATATTTTCGGAT	5517
QY	4379	AAATGTACTTAAAGTCTCAATTTAGTGAACTGAGACATTTGGCCCTTAGTAGGCTGTAT	4438	Db	26363	CCAGAGCCGATTGCTTTTAAAGCTAATAATGAATGTTTAAAGAGTACATATTTTCGGAT	26304
Db	27441	AAATGTACTTAAAGTCTCAATTTAGTGAACTGAGACATTTGGCCCTTAGTAGGCTGTAT	27382	QY	5518	CGTTCACTTAGTCAGCAATACACATATATAGCCCAATATTTTTTAAAAAATAAGAGCTT	5577
QY	4439	TCATCTGCTAATTTAAAAAGGGAGTACAGGATTTTATTAAGTAAAGCATTTTGGAAATGG	4498	Db	26303	CGTTCACTTAGTCAGCAATACACATATATAGCCCAATATTTTTTAAAAAATAAGAGCTT	26244
Db	27381	TCATCTGCTAATTTAAAAAGGGAGTACAGGATTTTATTAAGTAAAGCATTTTGGAAATGG	27322	QY	5578	GGTCAACCTCTATACACATATTAAGAGATATAGCCTTTTCAAAATGAATCTTAAACCT	5637
QY	4499	GGATAGCGCCATATATGATATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4558	Db	26243	GGTCAACCTCTATACACATATTAAGAGATATAGCCTTTTCAAAATGAATCTTAAACCT	26184
Db	27321	GGATAGCGCCATATATGATATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	27263	QY	5638	TTACAGAAACTTTTATAGGTTATGCTTTTATTTTAAAGACTTATTTTAACTCAAGTGC	5697
QY	4559	ACACACATACATCTTAAATCTTGCCCTGCATGAAATTCAAATACATGAGGCACATCTT	4618				

Db 26183 TTACAGAACTTTCTTATAGGTTAAGCCTTTATTTTAAAGCTTATTATTAATCAAGTGC 26124
QY 5698 CATTAGATGATATATATAGTGGCTTTGATATATAATGCTTTCTGACAAAATGTTAGA 5757
Db 26123 CATTAGATGATATATATAGTGGCTTTGATATATAATGCTTTCTGACAAAATGTTAGA 26064
QY 5758 TGGTATTTTAAACAGGTACATTTTACAGTGTGTTTCTTATCAATTTGCTATATTGACACAG 5817
Db 26063 TGGTATTTTAAACAGGTACATTTTACAGTGTGTTTCTTATCAATTTGCTATATTGACACAG 26004
QY 5818 AATCAGTGTGTTCTTTTCAATAGGTTTACAAAGTGTATTTTATTTTAAAGGTTTACGT 5877
Db 26003 AATCAGTGTGTTCTTTTCAATAGGTTTACAAAGTGTATTTTATTTTAAAGGTTTACGT 25944
QY 5878 GTCTCAAGCACACTGCTCTCCAGTACGTAAAGTTAAATAATACCAAGTTG 5937
Db 25943 GTCTCAAGCACACTGCTCTCCAGTACGTAAAGTTAAATAATACCAAGTTG 25884
QY 5938 CTTCTAGCCTACTGAGATCCATGTGACATTTGGAGGAGATCTTTTAAATGTTTAGTATTGCG 5997
Db 25883 CTTCTAGCCTACTGAGATCCATGTGACATTTGGAGGAGATCTTTTAAATGTTTAGTATTGCG 25824
QY 5998 TCATTAGCAATGGCTGGCTGTTAGTTCTGCTAAATGTTGCTTAAGTTGAAATTTGCTTGG 6057
Db 25823 TCATTAGCAATGGCTGGCTGTTAGTTCTGCTAAATGTTGCTTAAGTTGAAATTTGCTTGG 25764
QY 6058 TTTTCTCACACTGTGTCAGCCATGCTCAACACACAGATAAGTCTGTTGTGATCACA 6117
Db 25763 TTTTCTCACACTGTGTCAGCCATGCTCAACACACAGATAAGTCTGTTGTGATCACA 25704
QY 6118 TAGATCTACATAAGTTGTGAGTTTGTGCTTAAACCCATAGGGAGCTCTTTTGGGATC 6177
Db 25703 TAGATCTACATAAGTTGTGAGTTTGTGCTTAAACCCATAGGGAGCTCTTTTGGGATC 25644
QY 6178 ATAGAAAAGAGATCATGCAACAGCAATTTGGTGAAGGCACACTCAGATTGCACTTAGGGC 6237
Db 25643 ATAGAAAAGAGATCATGCAACAGCAATTTGGTGAAGGCACACTCAGATTGCACTTAGGGC 25584
QY 6238 CTTTCTATGATGTGTCAACCTCTGAGGATGGAAGGCAGTCTTTTATGATTTATCTAG 6297
Db 25583 CTTTCTATGATGTGTCAACCTCTGAGGATGGAAGGCAGTCTTTTATGATTTATCTAG 25524
QY 6298 CCTAGAAATGACACAGAACTATTGCTTAATGTATATAAAACACTTCATTATATAAGCTTCAGT 6357
Db 25523 CCTAGAAATGACACAGAACTATTGCTTAATGTATATAAAACACTTCATTATATAAGCTTCAGT 25464
QY 6358 GGTACAGATGAACACAGAAATGTTATCTCTCAGAAAACACTCTCTTCAATATTATATT 6417
Db 25463 GGTACAGATGAACACAGAAATGTTATCTCTCAGAAAACACTCTCTTCAATATTATATT 25404
QY 6418 GGATCATGCTGTAATGTAACTTGGGCTACACTCTTCTATGCTGTCTACAACTTCTCTGT 6477
Db 25403 GGATCATGCTGTAATGTAACTTGGGCTACACTCTTCTATGCTGTCTACAACTTCTCTGT 25344
QY 6478 CTCAITTCAGTCTGATTTTTTTTATCCATAGAAAAGCACTACATTAGGTGTAAGAGTGAC 6537
Db 25343 CTCAITTCAGTCTGATTTTTTTTATCCATAGAAAAGCACTACATTAGGTGTAAGAGTGAC 25284
QY 6538 AATATATTTTTTATCTGACTTAATTTGTCTAATAACAACTTTTACACCACCAATGTT 6597
Db 25283 AATATATTTTTTATCTGACTTAATTTGTCTAATAACAACTTTTACACCACCAATGTT 25224
QY 6598 ATTCATGTGCACTTCAAAAAGGAGATCTCGGACATGCAAAAGTTACAGAACCAACCCAG 6657
Db 25223 ATTCATGTGCACTTCAAAAAGGAGATCTCGGACATGCAAAAGTTACAGAACCAACCCAG 25164
QY 6658 CTTTGTGCCAAGGTGACTGTAACTCAGATGGAAGTGGGCTTTTATAATAGGTTGG 6717
Db 25163 CTTTGTGCCAAGGTGACTGTAACTCAGATGGAAGTGGGCTTTTATAATAGGTTGG 25104
QY 6718 AGTGAAGAACATGCTGTATGTTACTTAAACAGCCCTTTGAATTTTAAACAAATCTGGGAATCC 6777
Db 25103 AGTGAAGAACATGCTGTATGTTACTTAAACAGCCCTTTGAATTTTAAACAAATCTGGGAATCC 25044

QY 6778 ATTAGAAAACGATTCATCATACCTGGAACAATAGCTGGAAGCTGCTGAAATTTGATTTTTA 6837
Db 25043 ATTAGAAAACGATTCATCATACCTGGAACAATAGCTGGAAGCTGCTGAAATTTGATTTTTA 24984
QY 6838 GCTAATGAAAAGTGTTCGACTAGTACTTAAATGTTCTTAATGATAAAGTTTGGAGT 6897
Db 24983 GCTAATGAAAAGTGTTCGACTAGTACTTAAATGTTCTTAATGATAAAGTTTGGAGT 24924
QY 6898 CAAAATAGAAAAGAAAAAATCTGCAATTCAGGCCGAAATTTTGTATATTTTATTCGCAAT 6957
Db 24923 CAAAATAGAAAAGAAAAAATCTGCAATTCAGGCCGAAATTTTGTATATTTTATTCGCAAT 24864
QY 6958 TAAAATGCTATTCGTAATATTTGGGAAATCAAGTGGCTTATCATATATATCGTGTACTT 7017
Db 24863 TAAAATGCTATTCGTAATATTTGGGAAATCAAGTGGCTTATCATATATATCGTGTACTT 24804
QY 7018 AAAATGTAATTCACAACTACTGTTGTAATTTGTAATAAATATAGCAAAAGTATATTTT 7077
Db 24803 AAAATGTAATTCACAACTACTGTTGTAATTTGTAATAAATATAGCAAAAGTATATTTT 24744
QY 7078 TGTGTGTATTAAGCTCTGTAATAATAGCAATCACATTATGAAGCTGCAAGTATACAT 7137
Db 24743 TGTGTGTATTAAGCTCTGTAATAATAGCAATCACATTATGAAGCTGCAAGTATACAT 24684
QY 7138 TTTAAACATTACATCCAAAGAGCAGACTATTTATTTGCCATATACCAAGATTTAAAAATA 7197
Db 24683 TTTAAACATTACATCCAAAGAGCAGACTATTTATTTGCCATATACCAAGATTTAAAAATA 24624
QY 7198 TTAATTTGCTGCTAATAATAATAGTACTGAGCTTCTTGTGGCTACAGTGTATGTT 7257
Db 24623 TTAATTTGCTGCTAATAATAATAGTACTGAGCTTCTTGTGGCTACAGTGTATGTT 24564
QY 7258 TGCTTAAGAATAAGATATGTGAATTTCCACAAAATATATGAATAAAAT 7305
Db 24563 TGCTTAAGAATAAGATATGTGAATTTCCACAAAATATATGAATAAAAT 24516

RESULT 4
AC097274/c

LOCUS AC097274 182102 bp DNA linear HTG 18-JAN-2002
DEFINITION Mus musculus clone RP23-59B17, WORKING DRAFT SEQUENCE, 8 unordered pieces
ACCESSION AC097274
VERSION AC097274.5 GI:18201773
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 182102)
AUTHORS McCombie, W.R., Baker, J.P., Balija, V., Dedhia, N.N., de la Bastide, M., Kuit, K., King, L., Kirchoff, K.A., Miller, B., Nascimento, L.U., O'Shaughnessy, A.L., Preston, R.R., Santos, L., Shah, R.S., Spiegel, L.A., Palmer, L., Yang, C. and Zutavern, T.
TITLE Mouse Genomic Sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182102)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-2001) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
COMMENT On Jan 18, 2002 this sequence version replaced gi:16973738.
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory
Center code: CSHL
Web site: <http://www.cshl.org/genseq>
Contact: mccombie@cshl.org
----- Project Information
Center project name: RP23-59B17
Center clone name: RP23-59B17

NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 48370: contig of 48370 bp in length
48371 49082: gap of unknown length
49083 49531: contig of 4549 bp in length
95034 95745: gap of unknown length
95746 129089: contig of 33324 bp in length
129090 129781: gap of unknown length
129782 150082: contig of 20301 bp in length
150083 150793: gap of unknown length
150794 162439: contig of 11646 bp in length
162440 163150: gap of unknown length
163151 172039: contig of 8949 bp in length
172100 172810: gap of unknown length
172811 178596: contig of 5786 bp in length
178597 179307: gap of unknown length
179308 182102: contig of 2795 bp in length.

FEATURES

Location/Qualifiers
1..182102
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-59B17"

BASE COUNT 50661 a 36394 c 36240 g 53815 t 4992 others
ORIGIN

Query Match 38.6%; Score 2821.6; DB 2; Length 182102;
Best Local Similarity 73.1%; Pred. No. 0;
Matches 4933; Conservative 0; Mismatches 1294; Indels 525; Gaps 81;

QY 719 AAAGGCACATGGCTACTCATCTCTCCCTGAGAAACCCACACAGTCTTAATTATGAGAAAA 778
DB 60773 AAAGGCACATGGCTACTCATCTCTCCCTGAGAAACCCACACAGTCTTAATTATGAGAAAA 60714

QY 779 TGTTCACCGGAAAGATCATCTGAAGAATCACCTCCATACACACGACCCCTTAACAAAGAGA 838
DB 60713 TGTTCACCGGAAAGATCATCTGAAGAATCACCTCCATACACACGACCCCTTAACAAAGAGA 60654

QY 839 CGTTTAAGTCGGAAGATGTGGCAAGAACTACAAATACCAAGCTTGAATTTAAACGTCAC 898
DB 60653 CCTTTAAATCGGAAGATGTGGCAAGAGCTACAAATACCAAGCTTGGTTTAAAGCGACACT 60594

QY 899 TGGCCTTGATCGCGCAAGTGGTGACCTCACCTGTAAGGTATGTTTGCACAACTTTTG 958
DB 60593 TGGCCTTGATCGCGCAAGTGGTGACCTCACCTGTAAGGTATGTTTGCACAACTTTTG 60534

QY 959 AAAGCAGGAGTGTCTTCTGGAGCACCTTAAATCTCATGACGCAAGTCTGTTGGGGG 1018
DB 60533 AGAGCAGAGTGTCTTCTGGAGCACCTTAAATCTCATGACGCAAGTCTGTTGGGGG 60474

QY 1019 TTAAGGAAAAAGCAGCAGTGGGAACATTTGTGATGTCGCGGTCTACACCGGAAAGATG 1078
DB 60473 TGAAGGAAAAAGCAGCAGTGGGAACATTTGTGATGTCGCGGTCTACACCGGAAAGATG 60414

QY 1079 TCCGGACACATGTTGTCACACTGGAAGAAGGACTTCTCTGTGAGTATTGTGCAC 1138
DB 60413 TCCGGACACATGTTGTCACACTGGAAGAAGGACTTCTCTGTGAGTATTGTGCAC 60354

QY 1139 AGAGATTGGCGGAAGGATCACTGATCGACATATGAAGAAGATCAACAATCAAGAGC 1198
DB 60353 AGAGATTGGCGGAAGGATCACTGATCGACATATGAAGAAGATCAACAATCAAGAGC 60294

QY 1199 TTCTGAAGTCAAAACAGAACAGTGGATTTCTTGACCCATTTACCTGCAATGTGCTG 1258
DB 60293 TTCTGAAGTCAAAACAGAACAGTGGATTTCTTGACCCATTTACCTGCAATGTGCTG 60234

QY 1259 TGCCTATAAAGACGAGCTCTCTCGGTGATGCTCTTACCTTCCAGTGAACCTGTTATCAA 1318
DB 60233 TGCCTATAAAGATGAACCTCTCGGTGATGCTCTTACCTTCCAGTGAACCTGTTATCAA 60174

QY 1319 AGCCATTCAAAACACTTTGGAGTTAAACCTCTACACACTCCATTTCACTTCACTGAGTCA 1378
DB 60173 AGCCATTCAAAACACTTTGGAGTTAAACCTCTATAACTCTTATTAACACTCCATTTCACTGAGTCA 60114

QY 1379 GCTCGGATCTGCCACCAAAATGATCAACAACTTTACCTTTGGGAATGATGCCCAATAG 1438
DB 60113 GCTCTGGGTCTGCTACCAAAATGATCAACAACTTTACCTTTGGGAATGATGCCCAATAG 60054

QY 1439 ATATGACACACTGTTTCACTCCCTCTCACCACTTTCTTTTAAATATCGTTTCACTTACCT 1498
DB 60053 ATATGATGCTGTTTCACTCCCTCTCACCTTTTCAATATCGTTTCACTTACCT 59994

QY 1499 CATATGCAATTTCTATTTCTGAAAGAAAGACAGCCATTAAGGGGGGAAATGAGAGTTACC 1558
DB 59993 CATATGCAATTTCTATTTCTGAAAGAAAGACAGCCATTAAGGGGGGAAATTTGAGAGTTACC 59934

QY 1559 TGATGGAGTTACAAGGTGGGTGCTCTTATCCCAAGATTCTCAAGCATCTCTCATCAT 1618
DB 59933 TGATGGAGTTACAAGGTGGGTGCTCTTATCCCAAGATTCTCAAGCATCTCTCATCAT 1678

QY 1619 CTAAGCTAGGTTGGATCTCTCAGATTGGGTCCCTAGATGATGGTGCAGGAGACTCTCCC 1678
DB 59876 CTAAGCTAGGTTAGAGCTCTCAGATTGGGTGCTGCTGCGGAGACTCTCCC 59817

QY 1679 TATCCAAAAGCTCTATCTCATCAGTGAGCCCTTAAACACACACAGCATTTGGATTTTCTC 1738
DB 59816 TGTCAAAGAGCTCTATCTCTATCAGTGAGCCCTTAAACACACACAGCATTTGGATTTTCTC 59757

QY 1739 AGTTGTTTAAATTTATACCTTTAAATGGTCTCTCTTAAATCTCTATCATGCTGGGAGCC 1798
DB 59756 AGTTGTTTAAATTTATACCTTTAAATGGTCTCTCTTAAATCTCTATCATGCTGGGAGCC 59697

QY 1799 TTGGAATGAGTATTTCCAGGAGAAAGACACTTCTTCTGTTTCCAGCTCCCCACACAAA 1858
DB 59696 TTGGAATGAGTATTTCCAGGAGAAAGACACTTCTTCTGTTTCCAGCTCCCCACACAAA 59637

QY 1859 CACAGATCTTCAAGATCTCTGAAACACTATAGGCTTGGGTCTCTGCACTCATCTCAG 1918
DB 59636 CACAGATCTTCAAGATCTCTGAAACACTATAGGCTTGGGTCTCTGCACTCATCTCAG 59577

QY 1919 CAGCTTTTCAACAGCAGTTTAAAGCACAAGTACCCTCCACCTTTCCATCAAGCTTTTC 1978
DB 59576 CAGCTTTTCAACAGCAGTTTAAAGCACAAGTACCCTCCACCTTTCCATCAAGCTTTTC 59517

QY 1979 AGTAGGATTTCTGGGACATGGATTCAATTACAGAAATGATGTAGTGTGCTGCTAGATGA 2038
DB 59516 AGTAGGATTTCTGGGACATGGATTCAATTACAGAAATGATGTAGTGTGCTGCTAGATGA 59459

QY 2039 CCAATTTTATTTAGTGGCTAC-TTTTAAACAGATATAAAATTTCTGCTTTTGTATATA 2097
DB 59458 CCAATTTTATTTAGTGGCTAC-TTTTAAACAGATATAAAATTTCTGCTTTTGTATATA 59399

QY 2098 CAAATTTTCAATTAAGCCAGTATAAAATAGAAATAGCTTTTAAACTGAGCTTTTGGAA 2157
DB 59399 CAAATTTTCAATTAAGCCAGTATAAAATAGAAATAGCTTTTAAACTGAGCTTTTGGAA 2157

QY 2158 TTTGTGTTCAAGTTTAAAGTTTACCTGGGTATTTTGTGCTGATGCTGCTGCAATTTGCA 2217
DB 59388 TTTGTGTTCAAGTTTAAAGTTTACCTGGGTATTTTGTGCTGATGCTGCTGCAATTTGCA 2217

QY 2218 TTAAGACTTTTTTTTTTCCATATAGAAAGCCATTTATAGTAGTAACTTTTACAAATC 2277
DB 59284 TTAAGACTTTTTTTTTTCCATATAGAAAGCCATTTATAGTAGTAACTTTTACAAATC 59230

QY 2278 CCAATTTTCAATTTTAAATTTTAAATTTTCAATTTTGTCTTAAATCAAGTGGCTCT 2337
DB 59229 CCAATTTTCAATTTTAAATTTTAAATTTTCAATTTTGTCTTAAATCAAGTGGCTCT 59172

QY 2338 ACCTTTTGACATCTGGCTCATTTAAATAATTTAGCAATAGAAATTTGATATAAAG- 2396

Db	55118	AGTGTATTATGTGGCTTCCAAAGAGAGATCTTGATGTGCAATGTTTACAGAACAAA	55059
Qy	6653	CCCAGCTTTTGTCCACAAGTGACTGTAACTCAGAAATGAAAAGTGGCTTTTATAATAGGG	6712
Db	55058	CCCAGCTTTTGTCCACAAGTGACTGTAACTCAGAAATGAAAAGTGGCTTGA-CATAGGG	55000
Qy	6713	TGTGGAGTGAAGAACATGCTGTATGTACTTAACAGAGCCCTTTGAAATTTAACAAAAA	6772
Db	54999	TGTGGAGTGAAGAACATGCTGTAGAGTTACTTAACAGAGCCCTTTGAAATTTAACAAAAA	54940
Qy	6773	AATCCATTAG-GAAACGGATTGTCATCATACCTGAACATAAGCTG-----	6815
Db	54939	AATCCATTAGAGAAATGGATTCGATCATGCTGGAACATGAAGCTGACTGCTGTTTTTTT	54880
Qy	6816	-----GACTGCTGAAATTTGATATTTTATAGCTAATGAAAAAGTGT	6853
Db	54879	TTGTGTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATAGCTAATGAAAAAGTGT	54820
Qy	6854	TTGAGCTAGTACACTAAAATGTTCTTAATGATATAGATTTTAGTCAAAATAGAAAAGAAA	6913
Db	54819	TTGAGCTAGTACTCTAAAATGTTCTTAATGGTAAAAGTTTTTGAGTCAAAATAGAAATG-AA	54761
Qy	6914	AAAATCTGCATTCCAGCCCGAAATTTTGATATATTTTATTCGCAATTTAAAATATGCTATTCG	6973
Db	54760	AAAATCTGCATTCCAGACTGAAATTTGTATATTTTTTATTCGCAATTTAAAATATGCTATTCG	54701
Qy	6974	TAAATTTGGAAATCAAGTGGCTTATCATGTATATCGTGTACTTAAAATGTATTCACAAA	7033
Db	54700	TGATATTTGGAAAGCAAGTGGCTTATCATGTATATCATGTACTTAAAATGTATTCACAAA	54641
Qy	7034	CTACTCTGTATTGTATATAAATATAGACAAAGATCATATTTTTTGTCTGTGTATAGCT	7093
Db	54640	CTACTCTGTATTGTATATAAATATAGACAAAG-----TTATCTGTGTGTATAGCT	54588
Qy	7094	CTGTAAAATAGCAATCACATATAGAGCTGCAGTGATACTACATTTTAAACATTCACATC	7153
Db	54587	CTGTACAGTTGC-TTCACGCCATGAAACGTCAATAATCCCATAGCTGAGAAACATACAAC	54529
Qy	7154	CAAGAGACGAGACTATTTATTTGTCATATACAGATTTTAAAATATTAATTTGCTGCTAAT	7213
Db	54528	CAAAACAAGGAGGATGCTTTATTTGTCACACACCAAGCTGCACACAGTCACTTGTCTGC----	54473
Qy	7214	TAAATAATATAGTACTCGAGCTTCTTGTGGCTTACAGTGTATGTTTGTCTGTAGAAATAGA	7273
Db	54472	TGCGGCTTACCAACACAGCTGTTATGACCTGCACTGGGATTTGTTGT-AGGAAAAGA	54414
Qy	7274	TATGTAAATCCACAAAATATATGAATAAAAT	7305
Db	54413	TGGATGAATCTTGCGAATACTTGTAGTTGAAT	54382

[illegible]

QY	1139	AGAGATTTTCGGCGAAGGATCACCTGACCTCGACATATGAAGAGCTCACAATCAAGAGC	1198	QY	2218	TTAAGACTTTTTTTTTTCCATATAGGAAGCCATTTATTAGTAGTAACATTTTACAAATC	2277
Db	74786	AGAGATTTTGGAGAAAGGATCACCTCACCTCGACACATGAAGAGAGTCAATTCAGGAGC	74727	Db	73717	CTCAGAGTTTAAATTT-----CATATAGGAGCGCCATTTATTATCACTAAA-TTTTACAAGTC	73663
QY	1199	TTCTGAAGTCAAAAAGAGAGAGTGGATTTCTTCCCTGACCCATTTTACCTGCAATGTCTG	1258	QY	2278	CCATTTTCAAAATACCTTTTAGATCTTAAATTTTCAATTTTCTTAATTAACAGTGTCTCT	2337
Db	74726	TTCTGAAGTCAAAAAGAGAGAGTGGATTTCTTCCCTGACCCATTTTACCTGCAATGTCTG	74667	Db	73662	CTATGTTCAAAATACCTTTTAGATCTTAAATTTTCAATTTTCTTAATTAACAGTGTCTCT	73605
QY	1259	TGCTATAAAGAGAGAGTCTTCTCCGGTGAATGCTTCACTTCCAGTGAATGTTATCAA	1318	QY	2338	ACCTTTTGACATCTGGCTCAATTAATAAATTTAGCAATAGAAATGTAATTTGTATAAAG-	2396
Db	74666	TGCTATAAAGAGAGAGTCTTCTCCGGTGAATGCTTCACTTCCAGTGAATGTTATCAA	74607	Db	73604	-----ATTGACCTTTTTTTTTTTTGCATAATTTCACTAGTGAATCTTATGAAGTGT	73549
QY	1319	AGCCATTTCAACAACCTTTGCGAGTAACTTACCAACACTCCATTTCACTGTCATCAGA	1378	QY	2397	--TTTGTGAATACTCAAGGTTTAA-ATTTCCTTAAGCTTCTTAATTTGAATTAAT	2453
Db	74606	AGCCATTTCAACAACCTTTGCGAGTAACTTACCAACACTCCATTTCACTGTCATCAGA	74547	Db	73548	GTTGTGTGAATAACTGAAGTGTTCACATTTTTTTTACTAGCCACTAAATGGATTCATAA	73489
QY	1379	GCTCGGAGTCTGCCACCAATGATCAACACTTTTACCTTTGGGAATGACATGCCCAATAG	1438	QY	2454	CAA--GTGCTTCAAAATGAATTAAGAGTCCAGTTTCGGAAGATAATAAATTTTGTGTAGAT	2511
Db	74546	GCTCTGGTCTGCTCAACCAATGATCAACACTTTTACCTTTGGGAATGACATGCCCAATAG	74487	Db	73488	TAAGTGTGTTTTAAATGAGTTAAGATCCAGTTTGGGAGAGTATAGTGGTTTTCTTAAT	73429
QY	1439	ATATGACACTGTTATCCCTCTCACACCTTTCTTCAAAATATCCGTTTCAGTTCTACCT	1498	QY	2512	ACACCAATAATTTACAGATCAGTATATTTCTGAAGACTCTCTGTGTCTGGCTTAAATATTTG	2571
Db	74486	ATATGATGCTGTTCACCCCTCTCATCATCTTCTTCAAAATGCCATTCAGTTCTACCT	74427	Db	73428	ATACCATCATTTACAGATCAATATATTTTGGAGACT--GCTATTCTCTGGCTTAAATATTTT	73370
QY	1499	CATATCAATTTCTATTCTTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1558	QY	2572	CCATCTTTTATTATGAGCCCTTTTAAAGAAACAAACCCCTAAACACAAAGCATCAGTATTTAT	2631
Db	74426	CATACGCAATCTCTATTCTTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	74367	Db	73369	-----AAGTTTTAGTATCTAT	73354
QY	1559	TGATGAGTTTACAGGTGGCTGCCCTCTTTCATCCCAAGATTTCTCAAGCATCTGCATCAT	1618	QY	2632	AGCAAAAAGAGACTCTGTTAGTGACATGGCATTTCTGTCTCACTTAATAGTTGGCCCTAA	2691
Db	74366	TGATGAGCTTACAGGTGGTGCACCATCTTTCATCCCGAGGATTTCTCCAGCATCG--TCAT	74310	Db	73353	--CTAAAGAAACCTTAGAGAGGTAGCAGGTGTGTGTGTAGT-----	73312
QY	1619	CTAAGCTAGGTTGGATCTCTCAGATTTGGTCCCTAGATGATGTTGCGAGGAGACTCTCC	1678	QY	2692	ATTAGTACACAGGATATTTTGTGTCTTTCATCTTCTTAAACATGCTATCTTTTTCATTTA	2751
Db	74309	CTAAGCTAGGTTAGAGCTCTCAGATGGTTCGCGAGATGATGTTGTTGGGAGACTCTCC	74250	Db	73311	ACTAGCCCTCAGCTGGTTTGGCGCATTTCAATCATCTTTCCATA-----	73267
QY	1679	TATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACAGCATTTGGATTTTCTC	1738	QY	2752	ATAATAGTAATAGTATGSCATTTGGGCTTTCAGAGTCGATATATAGGTAGATCTCTTT	2811
Db	74249	TGTCAAAGAGCTCTATCTATCAGTGACCCCTTAAACACACAGCATTTGGATTTTCTC	74190	Db	73266	-----AACAGTGTGGCATCGAAGTCATCAGGATCAGTGTAGAAAGTTTGTAT	73215
QY	1739	AGTTGTTTAAATTTCAATCTTAAATGGTCTCCCTATATCTCTATCAGTGGGAGCC	1798	QY	2812	AGTCTTTTCCACCTTTTCACATCCAGGGTGGGTCAAGTCAGCCAGCAGCAATTTATTTTCA	2871
Db	74189	AGTTGTTCAATTTCAATCTTAAATGGTCTCCCTATATCTCTATCAGTGGGAGCC	74130	Db	73214	AUGCCCTGTACCTTTTCATACCCCA--GGATGGCTTCAGTATACAGGTTCTCATTTTGC	73157
QY	1799	TTGGAATGAGCTATTTCCAGAGAGAGCAATTTCTCTGTTTCCAGCTCCACACACAAA	1858	QY	2872	TTGTTGGCCCAAGGTAGTCCATAATC--TAGAGCCATTTTGAACCTGCAGCCATGAGGTG	2930
Db	74129	TTGGATGAGCTATTTCCAGAGAGAGCACTCTCTCTCTCTCAGCTGCCACACAAA	74070	Db	73156	ATGTTGGACACAGTATTTCCAGAGCTTAGAGCCATTTGGGACATAGACAGTAATA	73097
QY	1859	CACAGGATCTCAGGATCTCGAAACACTATAGGGCTTGGTCTCTGCACTCACTGTGAG	1918	QY	2931	TGTTTATCCCAAGTGAATGACTCAGCC---TCTGTGGGTGACAGACTTCTTAAGCAGGA	2987
Db	74069	CACAGGATCTCAGGATCTCGAAACACTATAGGGCTTGGTCTCTGCACTCACTGTGAG	74010	Db	73096	TGTGMAACCTTACATGGCTGGGTGCTCAATCTGTGGGTGATAGACTCTCTCAACAGGA	73037
QY	1919	CAGCTTTTACCAGAGTTAAGCAACAGTACACCCCTCCACCGTTTCCATCAAGCTTTTC	1978	QY	2988	AGATAGCTGGAAGCACATGTTTACATTTTGGGAACCTTGTGTAGGATCATGGCCCTGTGA	3047
Db	74009	CAGCTTTACCAGAGCTCAGCTCAGCTCAAGCACTACTTTGCCCGTTTCCACAGGCGTTT	73950	Db	73036	AGATGGATGTAAGCA---GCCAAACCCAGAGACTTGTCCAGAGATTT---CCACTGTT	72984
QY	1979	AGTAGGATCTGGGACATGATTCATTTACAGAAATGATGTAGTGTGCTGCTGAGTGA	2038	QY	3048	GCCAGGTTTAAACCTGGAATTTTGAAGTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA	3107
Db	73949	AGTAGGATCTGGGACGAGTCA--CCGAGGTGCTTGTGTAGCCCTGCTGCAATGA	73892	Db	72983	GCCAGAGCTGAAACCTGAACTCTTGGTGCAGAGTAAAGCCTA---TCACATCACTT	72928
QY	2039	CCATTTTATTTTAGTCCCTTAC--TTTAAACAGTATAAATAATTTCTGCTTTTGTATAA	2097	QY	3108	CTTGCTGAATTTGATATGTTTTTCTTCCCTTAAAGTAAAGTAAAGTAAAGTAAAGTAA	3167
Db	73891	CCATTTATTTTAGTCCCTTACTTTTAAACATATACAAAACCTTCTGCTTTTGTATAA	73832	Db	72927	CTAAGCAGTCTTGCCCGGTGAGCTCTTGC-----	72898
QY	2098	CAAAATTTTCAATAGCCAGTATAAATAAGAACTAGCTTTTAAACTGAGCTTTGGAACCA	2157	QY	3168	CAGTCTTACTCCGATGTTATCTTCTGATTTCAATGTGAATCCATCTTCTTCTGCAATTT	3227
Db	73831	TCAGTGTATGAAGCCAGTAAATAATAGGAACCTAGCTCTTTCGATAAGCTTTGAAATCG	73772	Db	72897	-----TCAGTGGTATCCATCTTGGCTGCTGTTTT	72867
QY	2158	TTTGTGTTTCAAGTTTACCTGGGTATTTTGTCTGATTTCACTGCGCAATTTGACATTT	2217	QY	3228	TTGATGGAATTTGAAGTTAAATGCAATTTAGAAACCTACCTGATGAACCTACCAAGT	3287
Db	73771	TTCTGTTTAACTAGTTT-----GCTATTCTCTGATTTCTGCGAGTGTGATAGT	73718	Db	72866	TGGATGAGAACTTAACAATTAGACCATCAGAAAGCAACTTGGAGGATTT---CTACAGT	72810
				QY	3288	TTTAAAGTACTAGAAATATATACAGTAAATCCCACTTTTCATGCTCTCTGGGAATGAT	3347

Db 72809 TTTACACACCAAGGTT-----GTAGACCTACTTTTA-----CTCTGGGATCAGG 72761
QY 3348 AGGAGTATTGCAAAATAGTTGAGTTGTTAGAGGGTAAACAAAGTAAAGTAAACAAACCTTA 3407
Db 72760 AGGAATGTTAGAAACAAAGTTGAATCTATAG-TGGAACACAGAGAAATTAAGCAACACC 72702
QY 3408 TCTTGTTTAAACATGAAATTAACAAATGAGAAATATATATTACCTGGAATATTATAGGC 3467
Db 72701 AAACGCAATGAGTGACACCTAGATGGAATGCACCTGAGT-----GAGGTAGCGATGG 72648
QY 3468 TTTTCTCCTACATTAGACAAACCAACATAATCTTCTTTAAAGGCTCTAATTAATATATTTTCT 3527
Db 72647 ACTTCTCTTGTGACAGACCCAGGAATGCGATCGATGTTCCAT-GTATATCTTCT 72589
QY 3528 AAGGTGAGTTGGGACATTAACTTAAGAAACATATCTATTAAGACCTTGTAAACACCTTA 3587
Db 72588 AAGGTTATTTGTGGCATCAACTTATGAATGTATCTATTAAGACCTTGTAAACACCTTA 72529
QY 3588 TTTTAGGACCTTTCCTGTTGGGATGGGGCGAAGGTGGGAGTTTTAAAGAGATATAT 3647
Db 72528 TTTTGGGACCTTCTGTTGGGCGAGGGTGAA-----GGGAGTTTATAGAAGGCTATGT 72473
QY 3648 AFTCTCTTTAAAAAACAAGAAAGAAA-----ATATTTCTGAGCACTCAITTA 3695
Db 72472 ATCTCTTTTAAAAAGAAAGAAAGAAACATATATTTTAAATTTCTGAGCACTCAITTA 72413
QY 3696 GGCCTATAT--GGAACTCTTCTTTTGTAGGGCGAGTTATACATCGCAATGCAAT 3753
Db 72412 GCCCTGTATGGGAACTCTTTCCCTCTGTGAAGGCCAGTTATCACTGCATATTGCAAT 72353
QY 3754 GTTTACCAAGATTTCTAAATATGATGAGATTAAGTAAATTTGTAATGTAATGTAATGGGGGCTT 3813
Db 72352 GTTTACCATCATC-CTAAATGTAATACAGATTTACTGACGATATGCAATATTATACAAAT 72294
QY 3814 ATTTGGGAGTAGTAATTTGTTGAAATGTAATTTGTAATGTAATGTAATGGGGGCTT 3873
Db 72293 GCCTGGGAATAGCTAAATTTGAGGAGAGTGTAACTGTCTAT-----TTGAATGGGGCTT 72238
QY 3874 CAATATATATATATATACACACACACATGACACATGCGCACTTCATAGAAAT 3933
Db 72237 CACTCTATGT-----CATACACACACCATGAGTGCATTCGTAGAAAT 72198
QY 3934 CAAGTTGCTCTCGAAGGAGCTTTGGCTCTGATATTTATCATGCTCCTATATTTTTT 3993
Db 72197 CAGAGTTGCTCTCTAGAGGTTCAAG-----CTCCTGGTT 72163
QY 3994 TAATCTTTGGACAGTAGTTTTTATACCTTATGATTTTAAATTTTATATGAAAAATTACA 4053
Db 72162 CTAACCTTAGACAGTAGTTTTTATACCTTATGATTTTAAATTTTATGTAAGAAATACAT 72103
QY 4054 TTTATTTAAAAAGTGTGTTCCAAAGCATTAATAATATATATGTTTAAAGAAAGTACAT 4113
Db 72102 TTTATTTAAAA---GTTTTTCCAAAGGCATTAATAATATATATGTTTAAAGAAATACAT 72046
QY 4114 TTTTAAATTTTCAAACTGCTCTAGCTTTTCAATTAGAGATATTTTTTCTGAAGTAG 4173
Db 72045 GTTT-AACTTTCAAGCTGCTCTAAGCTATGAGTGGGAGTGCATCTCTGAAGTAGG 71987
QY 4174 GCTTTTCGCTCTGCTTCATCTGCTCTCTTCTGTTTCTATGAAACAGATGCTTTACCTTA 4233
Db 71986 GCTCTTAGCTCTGCTTCATGACCGCTCC-TTAGTTTCTATGAACACAGATCACTTACCTTA 71928
QY 4234 AATCTTTAGTGAATGATTAGTGTTCATATTTGCTTTTAAATCAACCATATAAAAGGAA---A 4290
Db 71927 ATGTTCTAGTTGGATGGTTAGTGTTCATATTTGCTTTTAAATCAACCTTAAAAAGGAGGG 71868
QY 4291 AAAATTTGTGACAGACACAAAT-----AGAAAACTATTTTAAATGAATCACAAT 4345
Db 71867 ACATTTGTGTGACAGACACAAATGAATAAATAAARCTATTTTATGTAGAGTTCAAGC 71808
QY 4346 AGCAAGTTGGAGCACTTATTTCTGTTTAAAAATGTACTTAAAGAGTCAATCAATTA 4405

Db 71807 ACCACATATGAAGCACTAGTTTATCTATCTAAATTTGCTTAAGAAG---TCACATTA 71751
QY 4406 GTGAACCTGAGACATTTGGCCTTAGTAGGCTGTA-----TTCACTGCTPAATTTAAAAAGGG 4460
Db 71750 ATGACCTGAGACGTTGGCCTTCATAGGCTGTATGTTTCACTGCTAATTTGAAAAAGGG 71691
QY 4461 AGTACCAAGATTTATTAAGTAAAGCATTTTGGAAATGGGAATAGCGCATATATGATG 4520
Db 71690 AGTACCAAGATATATGAATAAAGCATTTTG--AAATGGGAATGGCGCATATATGTTG 71633
QY 4521 TATGTTATGTTGTGTGT----- 4539
Db 71632 TGTGTAT 71573
QY 4540 -----GGTGTGTATATATATACACACATATATATATATATATATATATATATATATAT 4591
Db 71572 TGTGTGTATATGTTGTGTAT 71513
QY 4592 AAATTC---AAATATATGAGGCACATCTTCAGGCGCACAGTGTAAATTTTGGAGTC 4647
Db 71512 TTACCTCTGACGGAATCTTAGTACACATCTTCCAGGGAACCAAGTGTAAATTTAGAGTC 71453
QY 4648 TTAATTTTTCATG--TGTAACACCTCTTTGCTGTTCACCCCCAGG-CTTGAAATAACAC 4704
Db 71452 TTAATTTTCTGTATGTATACACCTCTTTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 71393
QY 4705 TTCAGAGTAAGAGGAAATTCAGCTAATTTGTTTTT-----TAAAAATGACTGTAGTC 4755
Db 71392 TTTCAAAGTAAGAGGGAAGAAATTTGTTTTTTGTTTTTTGTTTTTTGTTTTTTGTTTTTTG 71333
QY 4756 GTCACCTAAACCC-TTTTTGAGAGAAATTTCTATTAAGATGAGGACAGCTGCTTATTGCA 4814
Db 71332 GTCACCTAAACCCTTTTTTGATAGATTTTCTATTAATAAATTCAGGCAG-----CTTTTTTA 71279
QY 4815 ATTGCACATGTTCTAAACAAGGATGTAACACAGAAATTCGCTTTTTTTTCCCTAGAAAAAG 4874
Db 71278 ATTGCACATATTTTAAACAAGACGTAAACAGAAATTTGCTCCACTCTCCCGTAGAAAA 71219
QY 4875 AT---TGTTTGTCTATGTCAACTAGATATGATTAATAAATAAGTATTGCCAATGCTGTT 4931
Db 71218 TTTCTGCTACTGTTTATTTTGAAGTAGGACTGCTTAGCGATAGGCTTGGCCAAATACCCCTT 71159
QY 4932 TTTCTTCTCTAGTGCCAGAAATCATTTCTTGAATTTCTGTTAGTGTGCTTGTAGCTTGTG 4991
Db 71158 TTTCTTCTC-----CATGGTCTCTGGAAATTTCTGATAGTGTGCTTGTAGCATGGC 71112
QY 4992 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5051
Db 71111 TGAAGAAAGGATTAACATTTACTGATACAGTT-----TATTGTGGCAT 71066
QY 5052 TTGGCCTCAGACAGATATTGAACCTCATTCAGTTTCACCTTCCACATGTATGTACAACT 5111
Db 71065 TTGGACCTCAGCGAGATGCTGAGCCCATTCAGTTTTCAGTTTCCACATGTGCGATGCA-- 71008
QY 5112 TAGTCAACAAACACGGAAGTTGAGTGTGGAAGGATCTTGGCAGCTGTGAAGCAATGCTATC 5171
Db 71007 ACGCATGCATCTCGCGGAGCGTGAAGTGGGAGGCGCTGAGCCTTAAGCAATAATAATC 70948
QY 5172 CATTCATGTATACAGTACCTTTTATAGTTATTCGATCTGTTTAAACCTTTCATTTTAA 5231
Db 70947 CATTCATTTTACCAATATACCTTTTGTAGTTTATGCTCATAGCTTT---CATTTTAA 70891
QY 5232 TCTTATTACCAAGTTCAGTTTTTAAACCTTCAATTTGCTGCTGGCTGATTATGATCACTC 5291
Db 70890 ACCCTATTACTTGTTCAGTTTCTATTAATTTAAGTTGCTTGGTGGGCTTGCCACCCGTC 70831
QY 5292 TGTGTGCAACTTTTTT---ATTTCATTTAGTGTCTTCTTCAAGCTGTGT--ATTTTTGCCT 5347
Db 70830 TGTGTGAGCTTTTTTTTAAATTCAGCATTTCTTTAAAGCTGTGTGGTTTTTTTGTGCT 70771
QY 5348 ATTTGTTGCTGTGCTTTTATTTTCTTAGTCAATTTGTGGAATATAGTATATATGCTGT 5407
Db 70770 ATTTGTTGCTGTGCTTTTATTTTC-TAGTCATTTTGTGGAATATAGTATATGCTGT 70712

Qy	5408	AA	TTTGGACAGTAGCGGTTTTTAAAAACCATATACTGACTGAAACATGAGCCAGAGCCGA	5467
Db	70711	AA	TTTGGACAGTAGCGGTTTTTAAAAACCATATACTGACTGAAACATGAGCCAGAGCTGA	70652
Qy	5468	TTG	CTTTATTAAAGCTAAATAAGATGTTAAAGAGTACATATTTTTCAGAGTCGTTCACTTA	5527
Db	70651	TTG	CTTTATTAAAGCTAAATAAGATGTTTATAGAGTACATA-TTTTCAGGATCAITTCATCTA	70593
Qy	5528	GTG	AGCAATACACATATTATAGGCCAAATATTTTTTAAAAAATAGAGCTTCGGTCAACCTC	5587
Db	70592	GTG	AGCAGTACACA---TATAGCCAA-----TTTAAAAATAGAGCTTGGTCAACCTC	70543
Qy	5588	TAT	ACTACACATATTTACAAGATATAGCACTTTCAAATGAATCTATAACCTTTACAGAAAC	5647
Db	70542	TAT	ACTTCACATACTACAAGATATAGCACTTTCAAATGAACCTATAATTTCTACAGAAAC	70483
Qy	5648	TTT	CTTTATAGTTATGCCCTTTTA-TTTTAAGACTTATTAT-----AATTCAGTGGCCATTA	5702
Db	70482	TTT	CTCTATAGTTATGCCCTTTTATTTTAAAGACTTATTATATPTCATTTTCAAAGTGCCATTA	70423
Qy	5703	GAT	GATATATATGTAGGCCCTTTGATATAAATGCTTTGTGTACA--AAAATGGTAGATGG	5760
Db	70422	GAT	GATATATATGTAGGCCCTTTGATATAAATCTTTGTGTACAAAAAAATGGTAGATGA	70363
Qy	5761	TAT	TTTAAACAGGTACATTTTTTACAGTGTTTTCTTATCAATTTGCTATATTGTCACAGAAT	5820
Db	70362	TAT	TTTAAACAGGTACATTTTTTACAGTGTTTTCTTACCAGTTTGCTATATTGTCACAGAAT	70303
Qy	5821	CAG	TGTGTCTTTTCATAAGGTTTTTACAATGGTTTATTTTTTTCACAGGTTTACGTCTC	5880
Db	70302	CAG	TGTGTGTCTTTTTCATAAGGTTTTTACAATGGGTTATTTTTTTCACAGGTTTACGTATC	70243
Qy	5881	TC	AAAGCACACTGTCTTCCAG----TAGCTAAGTTAAAAAATACCACTTCACCCAAAGTT	5936
Db	70242	TC	AAACACACTGTCTTCCAGTCCATAAGTAAGTTTAAAAATACCAGTTTCAACCCAGTT	70183
Qy	5937	GCT	TCTAGCCTACTGAGATCCATGTGCATTTGGAGAGAGATCTTTTAAATGTTTAGTATTC	5996
Db	70182	GCT	TCTAGCCTACTGAGATCTCTGTGCATTTGAAAGAGATCTTTTTCATATGTTAGTATTC	70123
Qy	5997	GT	CATTAGCAATCGCTGGCTGTAGTTCTGTGTAATGTGTGCCTAAGTTGAATTTGCTTT	6056
Db	70122	AC	CTTAAAGGCGAGACTGTTTTTTCTGTCAAAATGAGTTCTTTGTGTAGTTTGTCTC	70063
Qy	6057	GTT	TTTTCTCACACTGTGTCCAGAGCCATGCTTACAACACAGATAAGTCTGTGTGTATCCAC	6116
Db	70062	AT	TTTATTCATGC--TGTCCAGAGCCAGTCTTACAATCAATGTAGCTCGATTTTGTG---	70008
Qy	6117	AT	GATCTCATTAAGTTGTGCAGTTTTGTGCTTAAAAACCATAGGGAGCTCTCTTTGGGAT	6176
Db	70007	-TA	GCCTACATATAATACCTAAATTTGTTGT-----ACTCACAGAGCACTCTCTTTGAGAT	69953
Qy	6177	CAT	AGAAAAAGAGATCATGCAACAGCATTTGTGTGAGGCACACTCAGATTGCACTTAGGG	6236
Db	69952	TAT	GCAAGATATGTCTTCGTAGCAGACACAGATCACCATTAA-----G	69909
Qy	6237	CC	TTCTTATGATTTGTCAACCTCTGAGGATGGAAGGCAGTGTCTTTTGTATTTATCTA	6296
Db	69908	CC	TTTCTCAATGTTAGCAGCCC-CTGGGGATGGAAGGCAGTGTCTTTTGTATGAGTTATC	69850
Qy	6297	G---	CCTAGAAATGACACAGAACTATTGCTAATGTATATAAAACACTTCAATTATATAGCTT	6353
Db	69849	GAT	CTCTTAAAGATGGCGAGGAATTTGTACTGATGTATAAGACATTTCAATTACATAAGCTT	69790
Qy	6354	CAG	TGTACAGATGAACACAGAAATGAATGTTTATCTTCTCAGAAAAACACTCCCTCAATATTA	6413
Db	69789	CAG	TGTACAGATGAACACAGAAATGAATGTTCTGTCTTTCAGAAAAATCCCTCTTCAACATTA	69730
Qy	6414	TAT	TGGATTCATGCTGTCAATGTAACCTTTGGGCTACAACTCTTCAATGGTGTACAAACTTCT	6473
Db	69729	TAT	GGGATTCATGCTGTACTGTCACTTTGGGATACAACTCTTCAATGGTGTCTTAAACT--T	69672

Qy	6474	CTGCTCATTCAGTGGTATTTTTTTTATTCATAGAAAAGGACTACATAGTGTGTAAAAGT	6533
Db	69671	CTGCTCATTCAGTGTGTATTTTTTTTATTCATAGAAAAGGACTATATATAGTGTGTAAAAGGT	69612
Qy	6534	GTACAAATATATTTTTTATACCTGTGAC-TTAAATTTGTCACTTAAACAACTTTTACACCAACCAC	6592
Db	69611	GTACAAATATATTTTTTATACCTGTGACTTTAAATTTGTCATTAGCNAACTTTTACACCAACCAC	69552
Qy	6593	AATGTATTATGTGACACTTTGCAAAAGGAGATCTCGGACATGCAAAATGTTTACAGAACAAA	6652
Db	69551	AGTGTATTATGTGCGCTTGC AAAAGGAGATCTTCGATGTGCAAAATGTTTACAGAACAAA	69492
Qy	6653	CCGAGCTTTTGTCCCAAGGTCAGCTGTACTCAGAAATGGAAGTGGGCTTTTATATATAGGG	6712
Db	69491	CCGAGCTTTTGTCCCAAAAGTGACTGTAACTCAGAAATGGAAGTGGGCTTTGA-CAATAGGG	69433
Qy	6713	TGTGGAGTGAAGAACATGCTGTATGTTACTTAAACAGCCCTTTTGAATTTTAAACAAAACCTGGG	6772
Db	69432	TGTGGAGTGAAGAACATGCTGTGAGTTACTTAAACAGCCCTTTTGAATTTTAAACAAAACCTGGG	69373
Qy	6773	AATCCATTTAG-GAAACGGATTGGATCATCTGAAACATAAGCTG-----	6815
Db	69372	AATCCATTTAGAGAAATGGATTCATCATGCTGCTGAAACATAAGCTGCTGCTTTTTTTTGTG	69313
Qy	6816	-----GACTGCTGAAATTTGTTATTTTACCTAATGAAAAAGTGT	6853
Db	69312	TTGCTGTGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACTTAATGAAAAAGTGT	69253
Qy	6854	TTGGACTAGTACTCTAAAAATGTTCTAATGATAAAGTTTGGTCAAAATAGAAAAAGAAA	6913
Db	69252	TTGGACTAGTACTCTAAAAATGTTCTAATGGTAAAAAGTTTTTGGTCAAAATAGAAATG-AA	69194
Qy	6914	AAATCTGATCTCCAGCCGAAATTTGTATATTTTTTATTTGCAATTTAAATTTGCTATTCTG	6973
Db	69193	AAATCTGCATTTCCAGCTGAAATTTGTATATTTTTTATTTGCAATTTAAATTTGCTATTCTG	69134
Qy	6974	TAATATTGGGAAATCAAGTGGCTTATCATGTATATCGTGTACTTAAAAATGTTTACAAA	7033
Db	69133	TGATATTGGGAAAGCAAGTGGCTTATCATGTATATCATGTACTTAAAAATGTTTACAAA	69074
Qy	7034	CTACTGTTGTATTTGTATAAAATATAGACAAAGATCATATTTTTTGTGTGTGTATAAGCT	7093
Db	69073	CTACTGTTGTATTTGTATAAAATATAGACAAAGA-----TTATCTGTGTGTGTAAAGCT	69021
Qy	7094	CTGTAAAAATAGCAATCACATTTAGAGCTGCAAGTGATCACTACATTTTAAAAATTCACATC	7153
Db	69020	CTGTACAGTTGC-TTCACGCCATGAAACGTCAATANTCCATAGCTGAGAAACATACAC	68962
Qy	7154	CAAAAGACGAGACTATTTATTTGTCATATACCAGATTTTAAATATTAATTTTGTCTGCTAAT	7213
Db	68961	CAAAACAGGAGGATGCTTATTGTCCACACACCAAGCTGCACACAGTCACTTTGCTGC----	68906
Qy	7214	TAAATAATAGTACTGAGCTTCTTGTGGCTACAGTGTATTGTTTGTGTGAAGAATAAGA	7273
Db	68905	TGGCGTTTACCAACACAGCTGCTTATAGCTCGAGTGGGATTTGTGT-ARGGAAAAAGA	68847
Qy	7274	TATGTGAATTTCCAAAAATATATGAATAAAAT	7305
Db	68846	TGGATGAATTTCTCGAAATCTTGAGTTGCAAT	68815

RESULT 6	AC129839	278377 bp	DNA linear	HTG 08-OCT-2002
AC129839	Rattus norvegicus clone CH230-1B22, ***			
LOCUS	6 unordered pieces.			
DEFINITION	Rattus norvegicus			
ACCESSION	AC129839			
VERSION	AC129839.3 GI:23195804			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			

REFERENCE
AUTHORS

Rattus.
1 (bases 1 to 278377)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbacia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homsai,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,U., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs.R.
Direct Submission
Unpublished
2 (bases 1 to 278377)
Worley,K.C.
Direct Submission

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (04-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 278377)
Worley,K.C.
Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:22538586.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUEG
Center clone name: CH230-1B22

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 220445 bases at least Q40
Consensus quality: 225552 bases at least Q30
Consensus quality: 228018 bases at least Q20
Estimated insert size: 248009; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 247916: contig of 247916 bp in length
247917 248016: gap of unknown length
248017 272115: contig of 24099 bp in length
272116 272215: gap of unknown length
272216 273589: contig of 1374 bp in length
273590 273689: gap of unknown length
273690 275235: contig of 1546 bp in length
275236 275335: gap of unknown length
275336 276509: contig of 1174 bp in length
276510 276609: gap of unknown length
276610 278377: contig of 1768 bp in length.

FEATURES
source

1..278377
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-1B22"

misc_feature

1..1312
/note="wgs contig"

misc_feature

201866..203657
/note="wgs contig"

misc_feature

246189..247916
/note="wgs contig"

BASE COUNT 67642 a 48158 c 47319 g 67344 t 47914 others
ORIGIN

Query Match

Best Local Similarity 37.8%; Score 2761.2; DB 2; Length 278377;
Matches 4848; Conservative 0; Mismatches 1343; Indels 471; Gaps 81;

QY	719	AAAGGCATGGCTACTCATTTCTCTGAGAAAACCCACAAAGTAAATTATTGTGAGAAA	778
Db	144185	AAAGGCATGGCTACTCATTTCTCTGAGAAAACCCACAAAGTAAATTATTGTGAGAAA	144244
QY	779	TGTTTACCGGAAGATCATCTGAAGATACCTCCATACACGACCTTAACAAGAGAGA	838
Db	144245	TGTTTACCGGAAGATCATCTGAAGATACCTCCATACACGACCTTAACAAGAGAGA	144304
QY	839	CGTTTAAAGTCGAGAGATGTGGCAAGACTACATACCAAGCTTGGATTAAACCTCACT	898
Db	144305	CGTTTAAAGTCGAGAGATGTGGCAAGACTACATACCAAGCTTGGATTAAACCTCACT	144364
QY	899	TGGCTTTCATGCCGCAACAAAGTGTGTCACCTCACTGTAAGTATGTTTGAACCTTTTG	958
Db	144365	TGGCTTTCATGCCGCAACAAAGTGTGTCACCTCACTGTAAGTATGTTTGAACCTTTTG	144424
QY	959	AAAGCACGGGAGTGTCTCTGGAGCACCTTAAATCTCATGAGGCAAGTGTCTGTGGGG	1018
Db	144425	AAAGCACGGGAGTGTCTCTGGAGCACCTTAAATCTCATGAGGCAAGTGTCTGTGGGG	144484
QY	1019	TTAAGAAAAAAGACCAAGTCGGAACATTTGTGATGCGCCGGTTTCTACACCCCAAGGATG	1078

Db 144485 TGAAGAGAAAAGCACCAGTGTGACACTGCGAGCGCAGGTTCTACACCCGGAAGATG 144544
QY 1079 TCGGAGACACATGTTGTCACACTGGAAGAAGCACTTCCTCTCTCAGTATGTGCAC 1138
Db 144545 TTGGGAGACACATGTTGTCACACGGAAGAAGGACTTCCTCTCTCAGTATGTGCAC 144604
QY 1139 AGAGATTGGCGAAGAGATCCTGACTCGACATATGAGAAGAGTCACAATCAAGAGC 1198
Db 144605 AGAGATTGGCGAAGAGATCCTGACTCGACATATGAGAAGAGTCACAATCAAGAGC 144664
QY 1199 TTCTGAAGTCAAAAACAGAACAGTGAATTCCTTGACCCATTTACTCGCAATGTCTCG 1258
Db 144665 TTCTGAAGTCAAAAACAGAACAGTGAATTCCTTGACCCATTTACTCGCAATGTCTCG 144724
QY 1259 TGCTATAAAGACGAGCTCTTCCGGTGAATGCTCTTACCTCCAGTGAATCTTTATCAA 1318
Db 144725 TGCTATAAAGATGAGTCTCTCGGCTCATGTCTTGCCTTCAGCGAATCTTTGTCAA 144784
QY 1319 AGCCATTCAAAACACTTTGACGTTAAACCTCTACAACTCCATTTCACTGTCATGACA 1378
Db 144785 AGCCATTCAAAACACTTTGACGTTAAACCTCTATAACACTCCATTTCACTGTCATGACA 144844
QY 1379 GCTCGGATCTGCCACCAATCATCAAACTTTTACCTTTGGGAATGACATGCCCAATAG 1438
Db 144845 GCTCTGGTCTACTCCAAATGATTAACAATTTACTTTGGGAATGACATGCCCAATAG 144904
QY 1439 ATATGACACTGTTCACTCCCTCTCACACCTTTCTTTCAAATATCCGTTCACTTACCT 1498
Db 144905 ATATGACACTGTTCACTCCCTCTCATCATCTTCTTCAAATGAGTTCAGTTTACCT 144964
QY 1499 CATATGCAATTTCTATTCCTGAAAGAACAGCCATTAAGGGGGAATTTGAGATTAAC 1558
Db 144965 CATATGCAATTTCTATTTCCCGAGAAAGACAGCCCTTTAAAGGGGGAATTTGAGATTAAC 145024
QY 1559 TGATGAGTTTACAAGGTGGGCTGCTCTTCACTCCAAAGATTTCTCAAGCATCTGTCATCAT 1618
Db 145025 TGATGAGTTTACAAGGTGGGCTGCTCTTCACTCCAAAGATTTCTCAAGCATCTGTCATCAT 145081
QY 1619 CTAAGCTAGGTTGGATCTCAGATTTGGGTCCTTAGATGATGTTGACAGGAGACTCTCC 1678
Db 145082 CTAAGTTAGGTTTAGAGCTTCAGATTTGGGTCCTTAGATGATGTTGACAGGAGACTCTCC 145141
QY 1679 TATCCAAAGCTCTATCTCATCAGTACGACCCCTTAACACACAGCATTTGGATTTCTC 1738
Db 145142 TGTCAAAGAGCTCTATCTCTATCAGTGACCCCTGAAACACACAGCATTTGGATTTCTC 145201
QY 1739 AGTTGTTAAATTTCAATCTTAAATGGTCTCTCCCTATAATCTCTATCAGTGGGAGCC 1798
Db 145202 AGTTGTTCAATTTCAACCAATTAATGGTCTCTCTTACATCTGCTCTCGTGGAGCC 145261
QY 1799 TTGGAATGAGCTATTTCCGAGAAAGACAATTTCTTCTGTTTCCAGCTCCCAACACAAA 1858
Db 145262 TTGGATGAGCTATTTCCAGGACAGGCGCAATTTCTGTTTCTCAGCTGCCACACAAA 145321
QY 1859 CACAGGATCTTCAGGATCTGAAACACTATAGGCTTGGGCTCTCTGCACTCACTGTCAG 1918
Db 145322 CACAGGATCTTCAGGATCTGAAACACTATAGGCTTGGGCTCTCTGCACTCACTGTCAG 145381
QY 1919 CAGCTTTCCAGCAGTTTAAAGCAACAGTACACCTCCCAAGCTTTCCATCAAGCTTTTC 1978
Db 145382 CAGCTTTCCAGCAGCTTAAGCTCAAGCACTACCTGCCCCGTTTCCAGGCTTTTC 145441
QY 1979 AGTAGGATTTGGGACATGATTCATTTACAGAAATGATGTAGTGTGGCCCTAGATGA 2038
Db 145442 AGTAGGATTTGGGACATGATTCATTTACAGAAATGATGTAGTGTGGCCCTAGATGA 145499
QY 2039 CCATTTTATTTAGTGGCTACTTTAAACAGATTAATAATTTCTGCTTTGTATATAC 2098
Db 145500 CCATTTTATTTAGTGGCTACTTTAAACAGATTAATAATTTCTGCTTTGTATATAT 145559
QY 2099 AAATTTTCAATTAAGCCAGTATAAATAGAACTAGCTTTTAAACTAGCTTTTGAACCAT 2158

Db 145560 CAGGTTTATGAAGCCAGT---AAATAGGAATTAGCCTCTTTGATAAGCTTTGAAATGGT 145616
QY 2159 TTGTGTTTCAAGTTAAAGTTTACCTGGGTAATTTTCTCTGATTCTACTGCAATTTGTACATTT 2218
Db 145617 TCGTGTTTAACTACGTTT-----GCTGTTTCTCTGATTTCTCTGCAAGTTGTCAATGTC 145670
QY 2219 TAAGACTTTTCTTTTTCATATAGGAAAGCCATTTATAGTAGTAACCTTTTACAAATCC 2278
Db 145671 TCGGAGTTTATTT-----CATAGAGGAGGCCATTTATCTATAA-TTTTACAAGTCC 145725
QY 2279 CATTTTCAAAATTAATTTTATAGTCTTTAAATTTTCTTACTAGCTTCTTAATGATTAATCAA- 2338
Db 145726 TGTGTTCAAAATTAATTTTATAGCCTTTAACTTTCACTTTTGTCTCTAATAGCAATGG----- 145780
QY 2339 CCTTTTGAATCTGCGCTCATTTAAATTTTAGCAATAGAAATGTAATTTGTAATAAGTT 2398
Db 145781 GCATTTGCACTTTTCTTTTTCGAAATGTGCGATCAATGTAATCTTCTATGAAATGTG 145840
QY 2399 TGTGAATAACTCAAGGGTTT-AAATTTTCTTACTAGCTTCTTAATGATTAATCAA- 2456
Db 145841 TGTGAAGAACTGAGTGTTCACATTTCTTACTAGCCACTAATGATTAATCAA- 145900
QY 2457 -GTGCTTCAAAATGAATTAAGAGTCCAGTTCGGAAGATTAATGATTTGTTAGATACAC 2515
Db 145901 TGTGTTTAAATGAGTTAAGAAATCCAGTTTGGGAGAGTATAGTGTTCCTTAATACAC 145960
QY 2516 CATAAATTCAGATCAGTATATTTCTGAAGACTCTCTGTTCTGCTGGCTAAATATTTGCCAT 2575
Db 145961 CATCAATTTGAGATCAATATATTTTGAAGACT-GCTATTTGTCTGGCTAAATTAAGAGTTT 146019
QY 2576 CTTTATTTATGAGCTTTAAGAAAAACAAACCCCTAAACACAAAGCATCAGTATTTATAGCA 2635
Db 146020 CAGTA-----TCTATTT 146031
QY 2636 AAAAGAGACTCTGTTAGGTGACATGTCATTTCTGTCATTAATAGTTGGCCCTTAATTA 2695
Db 146032 AATAGAAACCTAGAGATTTACGGGTGTTCTCTGTAG-----AACTA 146075
QY 2696 GTACACAGGATATTTGTCTGTGTTTCTCTTCAATGCTATCTTTTCAITTAATAA 2755
Db 146076 GCCTTCAGGCTGATTTGCCACATTTCAACCATCTTTCCGT----- 146115
QY 2756 TAGTAATAGTATGGGCTCTTCAGAGTTCGATATATAGTAGATCTCTTTAGTC 2815
Db 146116 ---TACAGTGTGGCTCGGAGTCTTCAGGATCAGTGTAGAAGT-GTTCTGTGTATG 146171
QY 2816 TTTTCCACCTTTCAATCCAAAGGGGTGGGTCAAAGTGCAGCAGCAATTTATTTTATGTT 2875
Db 146172 CCCTTCACCTTTTCATACCCCGAGGATGGCTTCA--CAATACAGCAATTTTACTTTGCATG 146229
QY 2876 TGGCCACAGGTTAGTCCATAATCTAGAGCCATTTGGAACCTGAGCCATCAGGTTGCTTT 2935
Db 146230 TTGGTCCAGTTAGTGAGAGTCTTTAGAACATTTGCGGACATAGACAGTAATATANGA 146289
QY 2936 ATCCCAAGTGGATTTGACTCAGCTCTGT--GGGTGACAGACTTCTTAAGCAGGAAGATAG 2993
Db 146290 AGCCAGAGTGGCTGGCTGCTGAATCTGCGGGTGTATAGCTCTCCACAGGAAGATG 146349
QY 2994 ACSTGAAGACATGTTTACATTTGGGAACCTTGTGTAGGGATCATGGCCCTGTAGCCAGG 3053
Db 146350 ATGTAAACCTACAGCCAAACCCAGCACTTGTCTAGAGA--ATGTCCACTGTTGCCAGA 146407
QY 3054 GTTAAACCTGAGCTTTTAAAGTAAGTAAAGCAATAGCGCTTATCATTTCTTGTCT 3113
Db 146408 ACTGAGAACTGGAATCTCTATCGCATCACTT----- 146437
QY 3114 GAATTTGATATGTTTTTCTTCTTCCCTTAAAGATCAAAAGCAGAAAAACAAACACAGTCC 3173
Db 146438 -----CTAAACACAGCTCT 146452
QY 3174 TACTCCGATGTTATCTTTCTGATTTCAATGTGAATCCATCTTTCTTCTGCAATATTTTGGAT 3233
Db 146453 TGCC-----GTGCGCCCTTGCTCAGTGCATATACATCTTTCTCCTCGGTTTGGATG 146506

QY 3234 GGAGAAATTGAAGTTAAATGCAATAGAAAACCTACCTGATGAACCTACCAAGGTTTTTAAG 3293
Db 146507 GAGAACTTAAATAAACAACATCAAAAGCACTTGACGGTT---CCACAGTTTACAA 146563
QY 3294 TGACTAGAAATATATACAGTAAATCCACCTTCATGCTCTCTGGGAAATGATAGAGT 3353
Db 146564 CCACAAAGTTG-----TTAAGAGCCACTTTTACACATCTCCAGGAGCAGAGGAAT 146618
QY 3354 ATTGCAATAAGTTGAGTTTGTAGAGGTAACAAGTAAAGTAAACCAAACTTCTTGG 3413
Db 146619 GTTAGAAATAGTTGAATTTATAGAGAAACAGAGGAAATAAGCAAAACCAAAATGAA 146678
QY 3414 TTAACATGAAAATAACAAATGAGAAATATATATATTCACCTGAATAATATAGGCTTTCC 3473
Db 146679 TACGAATGCCAACAGGATGGTCTCACTGA-----GCAAGGTAGCAATGGGCTTCC 146731
QY 3474 TCACATTAGACAACCAACATAATCTCTTAAAGGTCTAAATATATATTTTCTAAGGT 3533
Db 146732 TCTTGTGACAGCAGCCACCAAGAATGTCATCGGTAGCT-CCGCGTATATCTTCTAAGGT 146790
QY 3534 CAGTTGGGACATTAACCTTAGAACAATCTATTAAGCACTTCTTAACACCTTATTTAG 3593
Db 146791 TATTTGGGCTCAACTAGACACTTAC-----TAAGCACTTGTTAACACCTTATTTCTG 146846
QY 3594 GACCCCTTTCGTTGGGATGGGGCAAGGTTGGAGGTTTTTAAAGAGATATATCTCT 3653
Db 146847 ACCCTTCTGTGGGCA-----GGAGTTTATAGAGATATATCTCT 146893
QY 3654 TTAATAAAAAACAGAAAGAAA-----TATTTGACACTCAATAGCCCT--A 3701
Db 146894 TTAATAAAAAACAGAAAGAAAATATTTTAAATCTCTGACACTCGTTAGCCCTGGC 146953
QY 3702 TATGAAACTTCTTCTTCTTTTGTAGCGCCAGTTATCACTGCAAGTGCATGTTTACCA 3761
Db 146954 TGGGAGAGCTTCTTCTTGTGTAGGCGGAGTTATCACTGCAATATGTTATTTTACC- 147012
QY 3762 AGAATTTCTAAAAATGAGTGCAGATTACTCAATATATATATTTTAAATATTTGGGA 3821
Db 147013 --CATCACTTAAGCAATAACAGATTACTGACTATGATGATTTTCAATGCTGGGA 147070
QY 3822 GTAGTAAATTTGTTGAGAAATGTAATTTGTAATATGTAATGTTGGGGCTTCAATATAT 3881
Db 147071 GTAGTCTAA--ATTGAGAAGTGAATCTGTCT-----GTGAATGGGCTTTCACATAT 147123
QY 3882 ATATATAATACACACACACATGACACATACACATACCCGCTTCTAGAAATCAAGTTG 3941
Db 147124 GT-----CATACACAGCCTAGTGCACCTTCATAGAAATCAGAGCTG 147163
QY 3942 CTCTCTGAGGAGCTTTGGCTCTCTGATATTTTATCATGCTCTATATTTTAAATCCCT 4001
Db 147164 CTCCCTAGAGG-GCTCAAGCTCTGTTCT-----AACCT 147198
QY 4002 GGAGCAGTAGTTTTTATCTTATGATTTTAAATTTTATGAAATTTTACATTTTAA 4061
Db 147199 AGAGCAGTAGTTTTTATACGTATGTTTAAATTTTATGTAAGGAAATTTTAAATTTTAA 147257
QY 4062 AAAAGTGTCTCCAAAGCATTAAATATATATATGTTTAAAGGAGTACATTTTAAAT 4121
Db 147258 AAGAAAGTGTCTCCAAAGGCTTAAATATATATACGTATTAAGGAAATACATTTT--AAC 147315
QY 4122 TTTTCAAACTGCTCTCTAGCTTTTGTATTAGGAAATATTTTCTGAAAGTAGGCTTTTCG 4181
Db 147316 TCTTCAAGCTGCTCTGAGTGAGGTTGGAGGCGCACTCACTCAGAAAGTGGGCTCTTCG 147375
QY 4182 CTCTGCTTCACTTCTCTTCTTCTTATGTTTCTATGAAACAGATTTGCTTACCTAAATCTT-T 4240
Db 147376 CATGACCGC-----TTCCTTAGTTTCTATGAAACCGATCACGCTACCTTAAATGTTCT 147426
QY 4241 AGTTCAATGATTTAGTTTCAATATTTGCTTTTATACCA-----TATPAAAGGAAAAAAT 4296
Db 147427 AGTTGGATGATTTAGTTTCAATATTTGCTTTTATACCACTTTTAAAGGAGGAGCAATTT 147486

QY 4297 GGTGACAGACGACAAAT-----AGAAAACCTATTTTAAATAGAAATACCAAAATAGCAAG 4351
Db 147487 GGTGACAGACGACAAAGTAAATAAAAAACCTATTTATTTATGTAGAGTTTACAGACACACA 147546
QY 4352 TGTGGAAGCACTACTTCTTATTTCTGTTTAAATGTACTTTAAGAGTCACTCAAAATTTAGTGAAC 4411
Db 147547 TATGAAAGCACTAGTTTATTTCTATCTTAAAAATTTACTTTAAGAG---TCACATTTATGACC 147603
QY 4412 TGAGACATTTGGCTTTAGTAGGCTGTA-----TTCACTGTCTAAATTTAAAAAGGAGTACC 4466
Db 147604 TGAGACGTTGGCTTCTGAGCTGTATATATTTCACTGTCTAAATTTGAAAAAGGAGTACC 147663
QY 4467 AGCATTTATTAAGTAAAGCAATTTTGGAAATGGGGAATAGCGCATATATGTATGTATGTG 4526
Db 147664 AGCATATATGAAATAAAGCAATTTTCGCAA--TGGGAATGGCGCATATATGTGTGTGTG 147721
QY 4527 TATGTGTGTGTG-TGTGTGTGTATATATACACACACATACATCTTAAATCTTGGCCC 4585
Db 147722 TATATATATATATATATGTATGTATATATATATATATATATATATATTTTATCTTAAT 147781
QY 4586 TGCATGAATTTCAATFACATGGAGGCACATCTTTCAGGGCACCACTGTATAAAATTTTGGAG 4645
Db 147782 TCTTGCCAGCGCAGAACTCTAATACACATCTTCCAGGGTACCAGTGTATAAAGCTAGAG 147841
QY 4646 TCTTAATTT---TTTCATGTGTACACCTCTTTCCTGTTCACCCGCCAGACT-TGAATAAA 4701
Db 147842 TCTTACTTCTCTTGTGTATGACACCTCTTTCGCTGTGTGTCGCCCCACAGACTTGAATGA 147901
QY 4702 CACTTCAGAGTAAAGAGGAATTCAGCTAATTT-TGTTTTTAAATTCAGCTGTAGTGTGTCAC 4760
Db 147902 CATTTCAAGTAAAGAGGAATAAGCAATTTCTTTTCGCTTTATGACTGTCAGTGTGTCAC 147961
QY 4761 TAAACCTTTTTCAGAGAAATTTCTATTAAGATAGGCGAGACTCGCTTATTTGAAATGCA 4820
Db 147962 TAAACCTTTTGTAGATATTTTATATTAATAATCACTGAGCTTTTATTTTCAITGCA 148021
QY 4821 CAATGTTCTAAAGAGTGAACAGAAATTCGCTTTTTCCTTA-----GAAAAGATTT 4877
Db 148022 CAATGTTTAAAGAGAACCAACAGAAATTCGCAACCCCTACCCCATGGAATACTGT 148081
QY 4878 GTTTGTTCTTGTCAACTAGATATGATTAAATAAGATTAAGCAATTCCTGTTTCAAT 4937
Db 148082 CACTGTTCTGTGTAGTAGGACTGCTTAGCAATAGGCTTGTAGTGCCTTCTCAAT 148141
QY 4938 CTCTAGTGCACAGAAATCAATCTTCTGAAATTTCTGTTAGTGCCTTAGCTTGGTTAAAAA 4997
Db 148142 CTC-----CATGGTCTCGGATTTCTGATAGTGCCTTAGCTT----- 148179
QY 4998 AAAAAAAGGGAATTAACATTAATAA---AGTAGTTTAGAATTTG 5054
Db 148180 -----AAAAAGGAGGTTTAAACATTAACATAAAGTGTAGTAGTTG 148226
QY 5055 GGCCTCAGACAGATATTGAACCTCAATTCAGTTTTCACATTCACATGATGTACAGTTAG 5114
Db 148227 GACCTCAGACAGATGCTGGGCGCGCTTCATTTTCACATTCACATGATGT-----GTGCA 148282
QY 5115 GTCACAAACGGAAGTTGAGTGTGGAAGGATCTTGGCACTGTGAAGCAATGCTATCCAT 5174
Db 148283 GCTTCAGGAGTGTGAGCGCTTTTCAGGAAAGGCTGCGCACTGTGAGCAATATAGTCCAT 148342
QY 5175 TGATCTATACAGTAGTACCTTTATAGTTATCGATCACTG--TTAAACCTTTCATTTTAAAT 5232
Db 148343 TGATATTTACCAATAATACCACCTCTGTAGTTAATGCTCTTAAGCTTCGCTTTTAAAT 148402
QY 5233 CCTATTACCAAGTTCAGTTTTCCTTAAACCTTCAATTCCTGCTGATTAATGCACTCACT 5292
Db 148403 CCCGTTA-CTTGTGAGGTTTCTTAAATGTAAGTTGCTTCGAGATCATGACACAGTCT 148461
QY 5293 GTGTGCAATTTTTT-----ATTTCATTTAGTGTGTTTCTTTTCAAGCTGTGATTTT 5343
Db 148462 GTGTGCACTTTTTTCCCTTTTAAATTCAGTTGGGTTTCTTTCACAGCTGTGTGTGTTT 148521
QY 5344 CCTATTGTTGCTTGTGCTTTATTTTCTTAGTCAATTTTGTGGAATATATAGTATATATTG 5403

On Sep 18, 2002 this sequence version replaced gi:21671673. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

[illegible]

3231	3290	3291	3351	221275	3411	221215	3471	221162	3531	221103	3591	221048	221001	3651	221000	3701	220940	3759	220880	3819	220823	3879	220770	3939	220730	3999	220695	4059	220636	4119	220578	4179	220518	4239	220467	4294			
QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	
GATGAGAAATTTGAAGTTAAATGCAATAGAAAACCTAAGTAACTACCAACAAAGTTTTT ATGGAGAACTTTAAACACATCAAAAGCAACTGACAGTT---CCACAGTTTTT AAGTCACTAGAAATATATACAGTAAATCCCACTTTTCATGCATCTCTGGGAATCATAGG ACACCACCAAAGTTGTTA-----AGAGCCCACTTTTACACAICTCCAGGAGACAGGAGG AGTATTGCAAAATAGTTGAGTTTGTAGAGGTAAACAAAGTAAAGTAAACAAACCACTATCT AATGTTAGAAAATAGTTGGAATTTATAGAGAAACAGAGGAAAATAAAGCAACCAAAAT TGGTTAAATGAAATTAACAATTTAGAAATATATTTATTTCACTGAAATTAATATAGGCTTT GAATACGAATGACCAAGGATGTGTGCTCACTGA-----GCAAGGTAGCAATGGGCT TCCTCACATTAGACAACCAACATAATCTTTTAAAGGTCTAATTAATATATTTTCTTAAG TCCTCTTGTCAGACAGCCACCAAGATGTATCGGTAGCT---CCGCGTATATCTTCCCTAAG GGTCAAGTTGGGACATTAACCTTAAGAAACATATCTATTAAAGCACCTTGTTTAAACACCTTAATT TAGGACCCCTTTCCGTTGGGATGGGGCAAGGGTGGGAGGTTTTTGAAGAGTATATATC GGTATTGTCGACATCAACCTACGAGACTTA-----CTAAGCACCTGTTTAAACACCTTATTC TGGACCCCTCTGTGTGGGCA-----GGGAGTTTTTATAGAAGAGTATATATC TCTTTAAAAAAAACAGAAAAGAAAA-----TATTCTCGAGCACTCAATTAGCCCT TCTTTAAAAAAAAGAGAGAACTATTTTAAATCTCTGAGCACTCGTTAGGCCCT --ATATGAAACTCTTCTCTTTTGTAGGCCAGTTATCACTGCAGATTGCAATGTTTA GGCTGGGGAAGCTCTCTCCCTTGTGTAGGCCAGTTATCACTGCATTTGTTGTTA CCAAGAAATTTCTAAATAGTGCAGATTACTGAAATATAATACATTAATTTAAATATTTG CCCATCACTAAGCAATAA---CAGATTAATGATGATGATGATTTTATACATGCGCTG GGAGTAGTAATTTGTGAGAAATGTAAATTTGTAATGTAATGAAATGGGGCTTCAATA GGAGTAGTCTAA---ATTGAGAGTGTAACTGTCCT-----GTGAATGGGGCTTCACTA TATATATATATATACACACACACATGCACATACCGCACTTCATAGAAATCAAAG TATGT-----CATACACAGCGCTAGTGCACCTTCATAGAATCAGAG TTGCTCTCTGAAGGAGCTTTGGCTCCGATATTTTATCATGCTCCTATATTTTTTAAATC CTGCTCCCTAGAGG-GCTAAGCTCTCTGGTTCT-----AAC CTTGGAGCAGTAGTTTTTATCTATGTTATTTAAATTTTATATGAAAAATTTACATTTAT CCTAGACAGTAGTTTTTATACGATGATGTTTAAATTTTATTTGTAATAA-TACATTTTA TAAAGAGTGTGTCGAAAGGCATTAATAATATATATGTTTAAATAGGAAGTACATTTTAA TTAAGGAAAGTGTTCGAAGGCATTAATAATATATACGTTTAAATAGGAAATACATTTT-- AAATTTTTCAAACTGCTCCTAGCTTTTCATTTAGGAGATATTTTTTCTGAAAGTAGGCTTT AACTCTTCAAGTGTCTCCTGAGCTGAGGGTGGAGCGCACTCACTCAGAAAGTGGGCTCT TCGCTCTGCTTCATTAATCTCTTTAGTTTCTTATGAAACAGATTGCTTACCTAAATCT T-TAGTTGAATGATGATGTTTCAATATGCTTTTAAATCAACA-----TATAAAGGAAAAA TCTAGTTGGAGATGATGATGTTTCAATATGCTTTTAAATCAACCACCTTTTAAAGGAGGACA ATTGGTGACAGACACAAAT-----AGAAAACCTATTTTTTAAATAGAAATCAAAATAGC																																							

D	b	219372	TTTGGCCTATGTCGGCTGCTTTATTTTTC--TAGTCAATTTGTGGAATATAGTCAATATA	219314	Q	y	6465	CAAACTTCTCTGTCCTCATTTCAGTCGCTATTTTATTTTATCCATAGAAAAGGACATACATTAGG	6524
Q	y	5401	TTGTGTTAAATTTGGACAGTAGCGGTTTAAAAACCATATATCTGACTGAAACATGAGCCA	5460	D	b	218261	CAAACT--TCTGTCTCATTTCAGTCGCTATTTTACCCATAGAAAAGGACT--CTATAGG	218205
D	b	219313	TTGTGTTAAATTTGGACAGTAGCGGTTTGTGAACCATATATCTGACTGAAACATGAGCCA	219254	Q	y	6525	TGTAAGAAGCTGTACAATATATTTTATCTGTGACTTAAATTTGTCATTATAACAACTTTTAC	6584
Q	y	5461	GAGCCGATTTGCTTTTATTAAGCTAATAATGAATGTTTAAAGAGTACATATTTTCAGATCGT	5520	D	b	218204	TGTAAGAAGTGTACAATATATTTTATCTGTGACTTAAATTTGTCATTATAACAACTTTTAC	218145
D	b	219253	GAGCTGATTTGCTTTTATTAAGCTAATAATGAATGTTTATAGAGTACATA--TTTCAGGATCAT	219195	Q	y	6585	ACCACCAAAATGTATTTTCATGTGCATTGCAAAAGGAGATCTCGGACATGCAAAATGTTTAC	6644
Q	y	5521	TCATCTAGTGAGCAATACATATATAGGCCAATATTTTAAABAAATAGAGCTTGGT	5580	D	b	218144	ACCACCAACAGTGTACTCATGTGCGCTTGCAAAAGGAGATCTTGGATGTGCAAAATGTTTAC	218085
D	b	219194	TCATCTAGTGAGCAGTACACA--TATAGGCCAA--TTAATAAAAAAATAGAGCTTGGT	219142	Q	y	6645	AGAACAAACCCAGCTTTTGTCCCAAGAGTGACTGTAACTCAGAAATGGAAGTGGGCTTTA	6704
Q	y	5581	CAACCTCTATACTACATATTAACAAGATATAGCACTTTCAAAATGAATCAAACTTTA	5640	D	b	218084	AGAACAAACCCAGCTTTTGTCCCAAGAGTGACTGTAACTCAGAAATGGAAGTGGAGCTTGA	218025
D	b	219141	CAACCTCTATACTACATATTAACAAGATATAGCACTTTCAAAATGAATCAAACTTTA	219082	Q	y	6705	TAAATAGGCTGTGAGTGAAGAACATCTGTATGTTACTAAACAGCCCTTTGAAATTTAAACAA	6764
Q	y	5641	CAGAAACTTCTTATAGGTTATGCTTTTA--TTTAAAGCTTATTAT--AAATCAAGT	5695	D	b	218024	-CATAGGCTGTGAGTGAAGAACATCTGTATAGTTACTTAACAGCCCTTTGAAATTTAAACAA	217966
D	b	219081	CAGAAACTTCTTATAGGTTATGCTTTTATTTTAAAGCTTATTATTATTTCAATTTCAAGT	219022	Q	y	6765	AAACTCGGAATCCCAATTAG--GAAACGGATTTGCATCATCTCGTAACATAAGCTGAGCTGTG	6823
Q	y	5696	GCCATTAGATCAATATATAGTACGCTTTGATATATATGCTTTGTGTACAA--AATGCT	5754	D	b	217965	AAACTCGGAATCCCAATTAGAGAAATGGATTTGCATCATCTCGTAACATAAGCTGAGCTGTG	217906
D	b	219021	GCCATTAGACATATATCCGTAAGCTTTGATATATATGCTTTGTGTACAAAGAATAGT	218962	Q	y	6824	AAATTTGATTTTATAGCTTAATGAAAAAGTGTTTGGACTAGTACTCT---AAATAATGTTCTA	6880
Q	y	5755	AGATGGTATTTTAAACAGGTACATTTTACAGTGTCTTCTATCAATTTGCTATATGCA	5814	D	b	217905	AATATTTTATTTTATAGCTTAATGAAAAAGTGTTTGGACTAGTACTCTAATAAATGTTTGA	217846
D	b	218961	AGATGGTATTTTAAACAGGTACATTTTACAGTGTCTTCTATCAATTTGCTATATGCA	218902	Q	y	6881	ATGTATAAGTTTTTCAGTCAAAATAGAAAGAAAAAATCTGCATTTCCAGCCCGCAATTTTG	6940
Q	y	5815	CAGAATCAGTGTGTCTTTTCAATAGGTTTTCATAGGTTTTCATAGGTTTTCATAGGTTT	5874	D	b	217845	ATGTATAAGTTTTTCAGTCAAAATAGAAATGAAATGAAATGAAATGAAATGAAATGAAATG	6940
D	b	218901	CAGAATCAGTGTGTGTCCTTTTACAAAGTGTTCATAGGTTTTCATAGGTTTTCATAGGTTT	218842	Q	y	6941	TATATTTTATTTGCAATTTAAATTTGCTATTCTGTATATTTGGAAATCAAGTGGCTTATC	7000
Q	y	5875	CGTGTCTCAAGACACATGCTCTCCAG--TACGTAAGTTAAAAATACAGTTTCAAC	5930	D	b	217785	TATATTTTATTTGCTTTTAAATTTGCTATTCTGTATTTGGAAATCAAGTGGCTTATC	217726
D	b	218841	CGTATCTCAAGACACATGCTCTCCAGTCCCAATAGTAAGTTAAAAAGATACCAAGTTCA	218782	Q	y	7001	ATGTATATCGGTACTTAAATATGATTTTCAAACTACTGTTGTTATTTGTTATATAAATATAG	7060
Q	y	5931	CAAGTGTCTCTAGCTACTCAGATCCATGACATTTGGAGGAGATCTTTTAAATGTTTA	5990	D	b	217725	ATGTATATCATGTACTTAAATATGATTTTCAAACTACTGTTGTTATTTGTTATATAAATATAG	217666
D	b	218781	CAAGTGTCTCTAGCTACTCAGATCCATGACATTTGGAGGAGATCTTTTAAATGTTTA	218722	Q	y	7061	ACAAAGATCATATTTTGTGTGTATAGCTCTGTAAATATAGCAATCAATTTATGAAG	7120
Q	y	5991	GTATTCGTCAATTAGCAATGGCTGGTGTAGTCTCGTAAATGTGTGCCCTTAAGT--TGA	6047	D	b	217665	ACAAAGA-----TGGTGTGTGTGTGTGTCTCTGTACAGTAGC--TTACAGCAGGAA	217614
D	b	218721	ATATTCGCCCTTATGAAGGACAGCTGGTGTCTTCTGCACAAATGAGTTCGTGTGTGA	218662	Q	y	7121	CTGCAGTGATACTACATTTTAAACATTCACATCCAAAGAGCAGACTTATTTATGTCAT	7180
Q	y	6048	ATTGTGCTTTTCTCAGTGTGTGAGGAGGATGCTTACACACATAGTCTGT	6107	D	b	217613	CTTCAACATACATATTTTAGAACCTCATCCAGAGAGGAGGCTGCTCAATG--CCAG	217555
D	b	218661	ATTGTGCTCGTCTTACT--CATGATGTGAGGAGGAGGATGCTTACACACATAGCTTGA	218604	Q	y	7181	ATACCGAGTTTAAATATTTAATTTGCTGTCTTAAATTAATAATAGTACTGAGCTTCTTGTG	7240
Q	y	6108	TGTGATCATAGATCTACATAAGTTGTGAGTGTGCTTAAACCCATAGGAGCTC	6167	D	b	217554	ACTCCCACTACACATACTACTCTCCCACT-----TCTCACACACAGCTTCTTGGG	217505
D	b	218603	T--TGTGTAAGCTCGTAAATACCTTAAATTTGTAATGT-----ACCCACAGACATC	218552	Q	y	7241	GCCTACAGTGTATTTGCTGTGAAGATAAGATATCTGAATTCACAAATATATGAT	7300
Q	y	6168	CTTTGGGATCATAGAAAAGAGATCATGCAACCCAGCATTTGTTGAAGGACACATCAGATTG	6227	D	b	217504	ACCTGCGACTAGTATTTTGTGT--AAGGAAAAAGATTTATGAATGCCACAGATATTTGAAT	217446
D	b	218551	CTCTGAGATTAAGCAAGCTAAT-----GCCATCAGCTTCTATAGCGCAGGATCA	218500	Q	y	7301	AAAAAT 7305	
Q	y	6228	CACCTAGGCGCTTCTATGATGTTGTGCAACCTCTGAGGATGGAAGGAGTGTCTTTTGA	6287	D	b	217445	AGAAAT 217441	
D	b	218499	CCAATA--AGCCCTTTCTACAATGTTAGCAGCC--CTGGAGATGGAAGGACGTGTCTTTGA	218442	Q	y	7301	AAAAAT 7305	
Q	y	6288	T--GTTATCTAGC--CTAGAAATGACACAGAACTATGCTAATGTTATAAACACTTCATTA	6344	RESULT 8				
D	b	218441	TAAGTATATCTATCTATAGATGACACAGGAATATTTACTGATGTTATAGACACTTCAATG	218382	AF057366				
Q	y	6345	TATAAGCTTTCAGTGTGACAGATGAACCAAGATGAATGTTTATCTTCTCAGAAAACTCTCT	6404	LOCUS				
D	b	218381	TATAAGCTTTCAGTGTGACAGATGAACCAAGATGAATGTTTATCTTCTCAGAAAACTCTCT	218322	DEFINITION				
Q	y	6405	TCAATATATATGGAATCATGCTGCTAAATGTAATGTAATGTTGGCTTACAACTTCTCATGGTGCTA	6464	AF057366				
D	b	218321	TCAACCTTCTATGGGATCATGCTGCTACTGTCACTTGGGATACAACTTCTCATGGTGCTA	218262	Mus musculus				
					AF057366.1				
					GI:6650194				
					cds.				
					Mus musculus				
					Mus musculus (house mouse)				
					Eukaryota; Metazoa; Chordata; Craniata				
					Vertebrata; Euteleostomi;				
					linear				
					1575 bp				
					mRNA				
					zinc finger protein				
					PLAG1 (Plagl)				
					mRNA, complete				
					AF057366				
					AF057366				
					AF057366.1				
					GI:6650194				
					cds.				
					Mus musculus				
					Mus musculus (house mouse)				
					Eukaryota; Metazoa; Chordata; Craniata				
					Vertebrata; Euteleostomi;				
					linear				
					1575 bp				
					mRNA				
					zinc finger protein				
					PLAG1 (Plagl)				
					mRNA, complete				

1438 CTCTGCACTCAGTCTGGCAGCGTTCCACGACGAGCTGAGCTCAAGCACTTACCTTGCCCC 1497

1961 GTTTCATCAAGACTTTTTCAGTAGGATCTGGACATGGATTCATTACAGAAATGATGTG 2020

1498 GTTTCATCAAGCTTTTTCAGTAGAGTCTGGCGGAGGAGCTCA--CCCAGAGTGCTTGTG 1555

2021 TAGCTGTGCGCTAGATGACC 2040

1556 TAGCCCTGCTGCAATGACC 1575

AC090187 65748 bp DNA linear HTG 13-APR-2001

LOCUS Homo sapiens chromosome 8 clone RP11-140116 map 8, LOW-PASS

DEFINITION SEQUENCE SAMPLING.

AC090187

VERSION AC090187.3 GI:13621271

KEYWORDS HTG; HTGS PHASE0.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 65748)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 8, clone RP11-140116

REFERENCE Unpublished 2 (bases 1 to 65748)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barina, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glende, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, I., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sougne, Z., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Apr 13, 2001 this sequence version replaced gi:13273404. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L12184

Center clone name: 140_I_16

NOTE: This record contains 78 individual * sequencing reads that have not been assembled into * contigs. Runs of N are used to separate the reads * and the order in which they appear is completely * arbitrary. Low-pass sequence sampling is useful for * identifying clones that may be gene-rich and allows * overlap relationships among clones to be deduced. * However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

742: contig of 742 bp in length

842: gap of 100 bp

1608: contig of 766 bp in length

1708: gap of 100 bp

2435: contig of 727 bp in length

2535: gap of 100 bp

3238: contig of 703 bp in length

3338: gap of 100 bp

4054: contig of 716 bp in length

4154: gap of 100 bp

4896: contig of 742 bp in length

4897: gap of 100 bp

5757: contig of 761 bp in length

5857: gap of 100 bp

6616: contig of 759 bp in length

6716: gap of 100 bp

7489: contig of 773 bp in length

7589: gap of 100 bp

8354: contig of 765 bp in length

8454: gap of 100 bp

9217: contig of 763 bp in length

9317: gap of 100 bp

10039: contig of 722 bp in length

10139: gap of 100 bp

10868: contig of 729 bp in length

10968: gap of 100 bp

11700: contig of 732 bp in length

11800: gap of 100 bp

12553: contig of 753 bp in length

12653: gap of 100 bp

13439: contig of 786 bp in length

13539: gap of 100 bp

14304: contig of 765 bp in length

14404: gap of 100 bp

15152: contig of 748 bp in length

15252: gap of 100 bp

16016: contig of 764 bp in length

16116: gap of 100 bp

16862: contig of 746 bp in length

16962: gap of 100 bp

17685: contig of 723 bp in length

17785: gap of 100 bp

18512: contig of 727 bp in length

18612: gap of 100 bp

19341: contig of 729 bp in length

19441: gap of 100 bp

20175: contig of 734 bp in length

20275: gap of 100 bp

21014: contig of 739 bp in length

21114: gap of 100 bp

21879: contig of 765 bp in length

21979: gap of 100 bp

22736: contig of 757 bp in length

22836: gap of 100 bp

23616: contig of 780 bp in length

23716: gap of 100 bp

24415: contig of 699 bp in length

24515: gap of 100 bp

25285: contig of 770 bp in length

25385: gap of 100 bp

26159: contig of 774 bp in length

26259: gap of 100 bp

27032: contig of 773 bp in length

27132: gap of 100 bp

27852: contig of 720 bp in length

27952: gap of 100 bp

28686: contig of 734 bp in length

28786: gap of 100 bp

29508: contig of 722 bp in length

29608: gap of 100 bp

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Boguslavskiy, L., Bouckgalter, B., Brown, A.,
 Camarata, J., Campiano, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T.,
 Lehoczyk, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
 McPheeters, R., Meldram, J., Meneus, L., Mihova, T., Mienga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
 Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J.,
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Apr 13, 2001 this sequence version replaced gi:13273404.
 All repeats were identified using RepeatMasker:

Smith, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12184

Center clone name: 140_I_16

* NOTE: This record contains 78 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 1 742: contig of 742 bp in length
 * 743 842: gap of 100 bp
 * 843 1608: contig of 766 bp in length
 * 1609 1708: gap of 100 bp
 * 1709 2435: contig of 727 bp in length
 * 2436 2535: gap of 100 bp
 * 2536 3238: contig of 703 bp in length
 * 3239 3338: gap of 100 bp
 * 3339 4054: contig of 716 bp in length
 * 4055 4155: gap of 100 bp
 * 4156 4896: contig of 742 bp in length
 * 4897 4996: gap of 100 bp
 * 4997 5757: contig of 761 bp in length
 * 5758 5857: gap of 100 bp
 * 5858 6616: contig of 759 bp in length
 * 6617 6716: gap of 100 bp
 * 6717 7489: contig of 773 bp in length
 * 7490 7590: gap of 100 bp
 * 7591 8354: contig of 765 bp in length
 * 8355 8454: gap of 100 bp
 * 8455 9217: contig of 763 bp in length
 * 9218 9317: gap of 100 bp
 * 9318 10039: contig of 722 bp in length
 * 10040 10139: gap of 100 bp
 * 10140 10868: contig of 729 bp in length

* 10869 10968: gap of 100 bp
 * 10969 11700: contig of 732 bp in length
 * 11701 11800: gap of 100 bp
 * 11801 12553: contig of 753 bp in length
 * 12554 12553: gap of 100 bp
 * 12554 13439: contig of 786 bp in length
 * 12554 13440: gap of 100 bp
 * 13440 14304: contig of 765 bp in length
 * 14305 14404: gap of 100 bp
 * 14405 15152: contig of 748 bp in length
 * 15153 15252: gap of 100 bp
 * 15253 16016: contig of 764 bp in length
 * 16017 16117: gap of 100 bp
 * 16117 16862: contig of 746 bp in length
 * 16863 16963: contig of 723 bp in length
 * 16964 17885: gap of 100 bp
 * 17886 18512: contig of 727 bp in length
 * 17886 18612: gap of 100 bp
 * 18613 19341: contig of 729 bp in length
 * 19342 19441: gap of 100 bp
 * 19442 20175: contig of 734 bp in length
 * 20176 20275: gap of 100 bp
 * 20276 21014: contig of 739 bp in length
 * 21015 21115: gap of 100 bp
 * 21115 21879: contig of 765 bp in length
 * 21880 21979: gap of 100 bp
 * 21980 22736: contig of 757 bp in length
 * 22737 23616: contig of 780 bp in length
 * 23617 23716: gap of 100 bp
 * 23717 24415: contig of 699 bp in length
 * 24416 25285: contig of 770 bp in length
 * 25286 25386: gap of 100 bp
 * 25387 26159: contig of 774 bp in length
 * 26160 26259: gap of 100 bp
 * 26260 27032: contig of 773 bp in length
 * 27033 27132: gap of 100 bp
 * 27133 27852: contig of 720 bp in length
 * 27853 27953: gap of 100 bp
 * 27954 28786: contig of 734 bp in length
 * 28787 29508: contig of 722 bp in length
 * 29509 29608: gap of 100 bp
 * 29610 30359: contig of 751 bp in length
 * 30360 30459: gap of 100 bp
 * 30460 31206: contig of 747 bp in length
 * 31207 31306: gap of 100 bp
 * 31307 32050: contig of 744 bp in length
 * 32051 32150: gap of 100 bp
 * 32151 32912: contig of 762 bp in length
 * 32913 33012: gap of 100 bp
 * 33013 33739: contig of 727 bp in length
 * 33740 33839: gap of 100 bp
 * 33840 34580: contig of 741 bp in length
 * 34581 34680: gap of 100 bp
 * 34681 35373: contig of 693 bp in length
 * 35374 35473: gap of 100 bp
 * 35474 36191: contig of 718 bp in length
 * 36192 36291: gap of 100 bp
 * 36292 37022: contig of 731 bp in length
 * 37023 37122: gap of 100 bp
 * 37123 37874: contig of 752 bp in length
 * 37875 37974: gap of 100 bp
 * 37975 38731: contig of 757 bp in length
 * 38732 38832: gap of 100 bp
 * 38833 39605: contig of 774 bp in length
 * 39606 39706: gap of 100 bp
 * 39707 40466: contig of 760 bp in length
 * 40467 40566: gap of 100 bp
 * 40567 41271: contig of 706 bp in length
 * 41272 41371: gap of 100 bp

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L24504
 Center clone name: 1053_A_18

* NOTE: This record contains 62 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 693: contig of 692 bp in length
 792: gap of 100 bp
 793 1456: contig of 664 bp in length
 1457 1556: gap of 100 bp
 1557 2165: contig of 609 bp in length
 2166 2285: gap of 100 bp
 2286 2963: contig of 698 bp in length
 2964 3063: gap of 100 bp
 3064 3778: contig of 715 bp in length
 3779 4538: gap of 100 bp
 4539 4638: contig of 660 bp in length
 4639 5277: contig of 639 bp in length
 5278 5377: gap of 100 bp
 5378 6099: contig of 722 bp in length
 6100 6199: gap of 100 bp
 6199 6892: contig of 693 bp in length
 6893 6992: gap of 100 bp
 6993 7707: contig of 715 bp in length
 7708 7807: gap of 100 bp
 7808 8515: contig of 708 bp in length
 8516 9299: gap of 100 bp
 9299 9399: contig of 684 bp in length
 9399 9400 9399: gap of 100 bp
 10101: contig of 702 bp in length
 10202 10201: gap of 100 bp
 10202 10896: contig of 695 bp in length
 10897 10996: gap of 100 bp
 10997 11660: contig of 664 bp in length
 11661 11760: gap of 100 bp
 11761 12438: contig of 678 bp in length
 12439 12538: gap of 100 bp
 12539 13233: contig of 695 bp in length
 13234 13333: gap of 100 bp
 13334 14057: contig of 724 bp in length
 14058 14157: gap of 100 bp
 14158 14836: contig of 681 bp in length
 14839 14938: gap of 100 bp
 14939 15642: contig of 704 bp in length
 15643 15742: gap of 100 bp
 15743 16461: contig of 719 bp in length
 16462 16561: gap of 100 bp
 16562 17250: contig of 689 bp in length
 17251 17351: gap of 100 bp
 17351 18053: contig of 703 bp in length
 18054 18153: gap of 100 bp
 18154 18862: contig of 709 bp in length
 18863 18962: gap of 100 bp
 18963 19683: contig of 721 bp in length
 19684 19784: gap of 100 bp
 19784 20488: contig of 705 bp in length
 20489 20588: gap of 100 bp
 20589 21312: contig of 724 bp in length

21313 21412: gap of 100 bp
 21413 22101: contig of 689 bp in length
 22102 22201: gap of 100 bp
 22202 22905: contig of 704 bp in length
 22906 23005: gap of 100 bp
 23006 23707: contig of 702 bp in length
 23708 23807: gap of 100 bp
 23808 24498: contig of 691 bp in length
 24499 24598: gap of 100 bp
 24599 25104: contig of 706 bp in length
 25105 25404: gap of 100 bp
 25405 26090: contig of 686 bp in length
 26091 26190: gap of 100 bp
 26191 26895: contig of 705 bp in length
 26896 26995: gap of 100 bp
 26996 27704: contig of 709 bp in length
 27705 27804: gap of 100 bp
 27805 28495: contig of 691 bp in length
 28496 28595: gap of 100 bp
 28596 29294: contig of 699 bp in length
 29295 29394: gap of 100 bp
 29395 30081: contig of 687 bp in length
 30082 30181: gap of 100 bp
 30182 30859: contig of 678 bp in length
 30860 30959: gap of 100 bp
 30960 31677: contig of 718 bp in length
 31678 31777: gap of 100 bp
 31778 32492: contig of 715 bp in length
 32493 32592: gap of 100 bp
 32593 33300: contig of 708 bp in length
 33301 33400: gap of 100 bp
 33401 34097: contig of 697 bp in length
 34098 34197: gap of 100 bp
 34198 34978: contig of 681 bp in length
 34979 35692: contig of 714 bp in length
 35693 35792: gap of 100 bp
 35793 36515: contig of 723 bp in length
 36516 37333: contig of 718 bp in length
 37334 37433: gap of 100 bp
 37434 38135: contig of 702 bp in length
 38136 38235: gap of 100 bp
 38236 38935: contig of 700 bp in length
 38936 39035: gap of 100 bp
 39036 39740: contig of 705 bp in length
 39741 39840: gap of 100 bp
 39841 40557: contig of 717 bp in length
 40558 41353: contig of 696 bp in length
 41354 42181: contig of 728 bp in length
 42182 42281: gap of 100 bp
 42282 42942: contig of 661 bp in length
 42943 43042: gap of 100 bp
 43043 43772: contig of 730 bp in length
 43773 43872: gap of 100 bp
 43873 44532: contig of 660 bp in length
 44533 44632: gap of 100 bp
 44633 45343: contig of 711 bp in length
 45344 45444: gap of 100 bp
 45445 46167: contig of 724 bp in length
 46168 46267: gap of 100 bp
 46268 46986: contig of 719 bp in length
 46987 47086: gap of 100 bp
 47087 47788: contig of 702 bp in length
 47789 47888: gap of 100 bp
 47889 48549: contig of 661 bp in length
 48550 48649: gap of 100 bp
 48650 49357: contig of 708 bp in length.

FEATURES
 source
 Location/Qualifiers
 1. .49357
 /organism="Homo sapiens"

<pre> /mcl_type="genomic DNA" /db_xref="taxon:9606" /chromosome="8" /map="8" /clone="RP11-1053A18" /clone_lib="RP11-11 Human Male BAC" BASE COUNT 12503 a 9244 c 8714 g 12754 t 6142 others ORIGIN </pre>									
<pre> Query Match 8.8%; Score 645.4; DB 2; Length 49357; Best Local Similarity 99.8%; Pred. No. 4.6e-105; Matches 646; Conservative 0; Mismatches 1; Indels 0; Gaps 0; </pre>									
QY	5155	CTGTAAGCAATGCTATCCATTCGATGATATCAAGTACCTTTATAGTATCGATCACTGTTA	5214						
Db	30213	CTGTAAGCAATGCTATCCATTCGATGATATCAAGTACCTTTATAGTATCGATCACTGTTA	30272						
QY	5215	AAACTTTCAATTTAAATCCTATACCAAGTTCAGTCTTTTAAACTTCAATGTCCTGG	5274						
Db	30273	AAACTTTCAATTTAAATCCTATACCAAGTTCAGTCTTTTAAACTTCAATGTCCTGG	30332						
QY	5275	CTGATTATGCATCACCTCTGTGTGCAACTTTTATTTTCAATTTAGTGTCTTTCTTCAAGCTG	5334						
Db	30333	CTGATTATGCATCACCTCTGTGTGCAACTTTTATTTTCAATTTAGTGTCTTTCTTCAAGCTG	30392						
QY	5335	TGTAATTTTGGCCTATTTGTTGCTGTGCTTTATTTTCTTAGTCAATTTGTGGAATATAGT	5394						
Db	30393	TGTAATTTTGGCCTATTTGTTGCTGTGCTTTATTTTCTTAGTCAATTTGTGGAATATAGT	30452						
QY	5395	GATATATTGTTGTTAATTTGGACAGTAGCGGTTTTTAAACCATATATCTGATGAAACAT	5454						
Db	30453	GATATATTGTTGTTAATTTGGACAGTAGCGGTTTTTAAACCATATATCTGATGAAACAT	30512						
QY	5455	GAGCCAGAGCCGATTGCTTTTATTAAGCTAAATGAATGTTTAAAGAGTACATATTTTCAG	5514						
Db	30513	GAGCCAGAGCCGATTGCTTTTATTAAGCTAAATGAATGTTTAAAGAGTACATATTTTCAG	30572						
QY	5515	GATCGTTCACTTAGTGAGCAATACACATATATAGGCCAATATTTTTTAAAAAATAGAG	5574						
Db	30573	GATCGTTCACTTAGTGAGCAATACACATATATAGGCCAATATTTTTTAAAAAATAGAG	30632						
QY	5575	CTTGTCACACCTCTATATACACATATTACAGATATAGCATTCGCTTTCAAAATGAATCTAAA	5634						
Db	30633	CTTGTCACACCTCTATATACACATATTACAGATATAGCATTCGCTTTCAAAATGAATCTAAA	30692						
QY	5635	CTTTTACAGAAACTTCTTATAGGTTATGCGCTTTATTTTAAAGACTTATATATTTCAAG	5694						
Db	30693	CTTTTACAGAAACTTCTTATAGGTTATGCGCTTTATTTTAAAGACTTATATATTTCAAG	30752						
QY	5695	TGCCATTAGATGATATATATGATAGGCCCTTTTGATATATAATGCTTTGTGTCAAAAATGGT	5754						
Db	30753	TGCCATTAGATGATATATATGATAGGCCCTTTTGATATATAATGCTTTGTGTCAAAAATGGT	30812						
QY	5755	AGATGGTATTTTAAACAGGTACATTTTACAGTGTCTTTCTTATCAAT	5801						
Db	30813	AGATGGTATTTTAAACAGGTACATTTTACAGTGTCTTTCTTATCAAT	30859						
<pre> RESULT 12 G25280/c LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE Mapped STGs </pre>									

Unpublished (1995)

Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu

Primer A: AACCAAGCTTAAGGCACCTACCA
 Primer B: CACAGAATGGCTTTTTC
 STS Size: 150
 PCR Profile:
 Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35
 Thermal Cycler:
 Protocol:
 Template: 10 ng
 Primer: each 5 pM
 dNTPs: each 4 nM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

Derived from dbEST (genbank accession D59273).

FEATURES	source	Location/Qualifiers
		1..421
		/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/map="314.2 cR from top of Chr8 linkage group"
STS		1..150
primer_bind		1..21
primer_bind		144 a 77 c 75 g 122 t 3 others
BASE COUNT		
ORIGIN		
Query Match		5.7%; Score 416; DB 11; Length 421;
Best Local Similarity		99.5%; Pred. No. 7.3e-64;
Matches		416; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY		4575 AAATCTTGCCCTGCATGAATTCAAATACATGAGGCACATCTTCAGGCGACACAGTGTTA 4634
Db		418 AAANCTTGCCNGCATGAATTCAAATACATGAGGCACATCTTCAGGCGACACAGTGTTA 359
QY		4635 AAATTTGGAGTCTTAATTTTCATGTGTACACCTCTTTSCCTGTTCACACCCCCAGACTT 4694
Db		358 AAATTTGGAGTCTTAATTTTCATGTGTACACCTCTTTSCCTGTTCACACCCCCAGACTT 299
QY		4695 GAAATAACACTTCAGAGTAAGAGGAATTCAGCTAATTTGTTTTTAAATTCAGCTGACT 4754
Db		298 GAAATAACACTTCAGAGTAAGAGGAATTCAGCTAATTTGTTTTTAAATTCAGCTGACT 239
QY		4755 GGTCACTAAACCCCTTTTGGAGAGAATTTCTATTAAAGATGAGGCAGACTCGCTTATTGA 4814
Db		238 GGTCACTAAACCCCTTTTGGAGAGAATTTCTATTAAAGATGAGGCAGACTCGCTTATTGA 179
QY		4815 APTGCACAATGTTCTTAACAAGGATGTAACACAGAAATTCGCTTTTTTTTCCCTAGAAAAAG 4874
Db		178 APTGCACAATGTTCTTAACAAGGATGTAACACAGAAATTCGCTTTTTTTTCCCTAGAAAAAG 119
QY		4875 APTGTTTGTTCTATGTCAACTAGATATGANTAAAAAATGATTTGCCAATGCTGTTTC 4934
Db		118 APTGTTTGTTCTATGTCAACTAGATATGANTAAAAAATGATTTGCCAATGCTGTTTC 59

4935 ATTCTAGTGGCCAGATCATTTATCTTGAATTTCTGTAGTGCCTTAGCTTGGT 4992
|||||
58 ATTCTAGTGGCCAGATCATTTATCTTGAATTTCTGTAGTGCCTTAGCTTGGT 1
|||||

RESULT 13
LOCUS BC023655
DEFINITION Homo sapiens, clone MGC:29597 IMAGE:4764127, mRNA, complete cds.
ACCESSION BC023655
VERSION BC023655.1 GI:23271119
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1829)
Strausberg, R.
Direct Submission
Submitted (05-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs@mail.nih.gov
Tissue Procurement: Louis Staudt
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAL Plate: 39 Row: f Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6031195.
Location/Qualifiers
1. .1829
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:29597 IMAGE:4764127"
/tissue_type="Primary B-Cells from Tonsils"
/clone_lib="NIH MGC 48"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
213. .1703
/codon_start=1
/product="Unknown (protein for MGC:29597)"
/protein_id="AAH23655.1"
/db_xref="GI:23271120"
/translation="MTFTSPVPPVLPQDAKQEEVGVKLVPRGRREASOVKCCCEI
SGTFFSNGEKLRPHSLUPQEPREYPCQLHCGKAPASKYLYRHMATHSAQKPHQMY
CDXMFHRKDLNHLQTHDPNKEALHSCGKNYNTKGLYRRLHWAHSAASGLSKV
CLQTFSTQALLHLKHAHRVAGAKKKPCDCHDRFRYTRKDRVRLVHTGRK
FLCOYCAORFGRKDLHTRVKKSHOELLKITEPVMGLILSCSSTVSVKELSPVL
CMASRDVMGTFKAPGMLPMGYGAHTPTMPSTGVPHSLVHNTLPMGMSVPLESSPIS
PAQLPKYQLGTSYLPDLKPKVVDSPFLAELPGSLSLSAEPQAPASPOPAALLD
EALLAKSPANLSALCAANVDFSHLGLFLPLNLPNCNPPGATGGLVWGVSQAQPELL
TTLQAQPDQSPGAGGGLNFGPLHPLPFTVFTSGLSLSTTLPRFHOAQ"

CDS

BASE COUNT	399 a	577 c	493 g	360 t
ORIGIN				
Query Match	5.4%; Score 392.8; DB 9; Length 1829;			
Best Local Similarity	58.3%; Pred. No. 7.7e-60;			
Matches 826; Conservative	0; Mismatches 542; Indels 48; Gaps 6;			
QY	577	AACTTTCTTCGCAACTGTGTGACAAAGGCGCTTTAAACAGTGTGTGAGAAATTAAGGTTAC	636	
Db	327	AAAGTCCAATGTGAAATTTTCGGGAAACACCTTTCTCAATGGGAGAGAGCTGAGGCTCAC	386	
QY	637	TCCTACTCTCACACAGGAGAGGCGCTTACAGTGTGATACACAGAGACTGCACAGGCC	696	
Db	387	AGCCTCCCGCAACACAGAGAGAGCCATATAGCTGCGCTCAGCTGCACCTGTGGCAGGTT	446	
QY	697	TTTGTCTTCTAAGTACAAATTTACAAAGGCGACATGGGTACTCACTTCTCTCAGAAAAACCCAC	756	
Db	447	TTTGCTTTCCAAATACAGCTGTATAGGCACATGGCCACCCACTCAGCCGAGAAACCCAC	506	
QY	757	AAAGTGAATTAATTTGTGAGAAATGTTCACCGGAAAGATCATCTGAAAGAAATCACCTCCAT	816	
Db	507	CAGTGTATGTACTGTGATTAAGATGTTTACCGCAAGGACCATCTGCGGAACCATCTGCAG	566	
QY	817	ACACAGCACCTTAAACAGAGAGCTTTAAGTGGGAAGAAATGTGGCAAGACTTACAATACC	876	
Db	567	ACCATGATCTTAACAAAGAGGCGCTCCACTGCTCTGAGTGGGTAAAGAAATTAACAATACG	626	
QY	877	AAAGTTGGATTTAAACGTCACCTTTGGCTTGCATGCGCAACAAAGTGGTGCACCTCACCTGT	936	
Db	627	AAAGTGGGTACCGGCGCCACCTGGCCATGCACTGCGCAGACGCGGTGACCTCAGCTGC	686	
QY	937	AAAGTATGTTTGCACACTTTTGAAGACACGGAGTGTCTTCTGAGACACCTTAAATCTCAT	996	
Db	687	AAAGTGTGCTGCAGACCTTTGAGAGTACCAGGCGCTGTAGAGCACCTGAAGGCCAC	746	
QY	997	GAAGGCAAGTGTCTGTGGGTGTTAAAGAAAAAGACACAGTGCAGCAATTTGTGATCG	1056	
Db	747	TCACGCGCGGTGACAGCGCGTGCACAGGAGAGACACCCCTGTGACCACTCGCAGCCG	806	
QY	1057	CGGTTCTACACCCGAAAGATGTCCGAGACACATGGTGGTGACACTGCGAAGAAAGGAC	1116	
Db	807	CGGTTCTATCTCTGTAGAGATGTACGCGCGCACTTAGTGTGACACAGCGCGTAGGAC	866	
QY	1117	TTCTCTGTCTGATTTGTGCACAGAGATTTGGCGAAAGATCACTGTGACATATG	1176	
Db	867	TTCTCTGTGCGAGTACTGTGCCAGCGGTTTGGCGGTAAAGACACCACTGACGCTCATGTC	926	
QY	1177	AAGAAGACTCACATCAAGAGCTTCTGAAGGTCAAAACAGAACAGCTGGATTTCTTGTAC	1236	
Db	927	AAGAAGAGCCACTCGCAGGAGCTGTCTCAAGATCAAGACAGAGCGCCGTGGACATGTTAGGC	986	
QY	1237	CCATTTTACCTGCAATGTGTCTGTGCTTATAAAGACGAGCTCTCTCCGGTGTGCTCTTA	1296	
Db	987	CTACTCAGCTGAGCTCCACAGTGTGAGAGAGAGCTGAGCGCTGTGCTGTGCATG	1046	
QY	1297	CCTTCCAGTGA---ACTGTTATCAAGGCCATTCAAAAACATTTTGAGTTTAAACCTCTAC	1353	
Db	1047	GCCTCTCGGACGTAATGGGACCAAGGCGCTTCCCTGGCATGTTGCCCATGGGCGATAT	1106	
QY	1354	AACACTCCATTTTCACTCCATGAGAGCTGGGATCTGCCCA---CCAAATGATCACAACT	1410	
Db	1107	GGTGGCCACATCCCTACCATTGCCAGCAGCGGGGTGCCACACTCCCTGGTGTGACACAG	1166	
QY	1411	TTACCTTTGGGAATGATGCGCCCAATAGATATGACACACTGTTTCACTCCCTCTCACACCTT	1470	
Db	1167	CTGCCCATGGGTATGAGTACCTCTGTGATCTCTACCTATCTCTTCCCAGCTCAGCTC	1226	
QY	1471	TCCTTCAATATCCGTTTCAGTTCTACTCATATGCAATTTTCTATTTCTGTAAGAAACAG	1530	
Db	1227	CCTCCAAATACCAAGCTTGGATCTACCTCATCTTG-----CCGACAAA	1271	
QY	1531	CCATTAAAGGGGAAATTTGAGAGTTACCTGAGTTTACAGGTGCGGTGCGCTCTTCTCA	1590	

```
Db 1272 TTGCCAAAGTGGAGTGATAGTTTCTGCGGAGAGCTTCTCGAAGCCTCTCTCTCA 1331
Qy 1591 TCCCAAGATTCTCAAGCATCTCATCATCTAAGCTAGGTTGGATCTCTCAGATTGGTTC 1650
Db 1332 TCCGCTGAACCCAGCCGCTCA-----CCTCAGCCGGGCA 1370
Qy 1651 CTAGATGATGGTGAGAGACCTCTCCCTATCCAAAAGCTCTATCTCATCAGTGACCCC 1710
Db 1371 GCTCGGCCCTCTTAGATGAAGCACTGCTTCCCAAGAGCCCGCCAACTCTCTGAGGCC 1430
Qy 1711 CTAACAACACAGCATTGGATTTTCTCAGTTGTTAAATTCATACCTTTAAATGGTCT 1770
Db 1431 CTCTGCGCTGTAATGAGCTCTCCCACTCTGAGCTCTTCTTCCACTCAACCTGCC 1490
Qy 1771 CCTATATA-----TCTCTATCAGTGGGAGCCTTGGAAATGAGCTATCCAGGAGAAGCA 1827
Db 1491 CCGTGAACCCACCTGGGGCCACAGGAGGCTGTCTATGGCTACTCCAGGCTGAGGCA 1550
Qy 1828 CATTTCTCTGTTCCAGCTCCCAACACACAGAGATCTTCAGGATCTCTGCAAACT 1887
Db 1551 CAGCCCTGCTTACCACTTTGCAAGCTCAGCCTCAAGATTCCCAAGGAGCTGGGGACCA 1610
Qy 1888 ATAGGGCTTGGGCTCTGCACTCACTGTGAGCAGCTTTCCAGCAGTAAAGCAAGT 1947
Db 1611 CTGAACCTTGGGCTCTGCACTCTTGGCTCTCTTCACTGCTGCGCTGAGT---AGC 1667
Qy 1948 ACCACCTCCCAAGTTCATCAAGCTTTTCAGTAG 1983
Db 1668 ACCACCTGCTGCTTTCATCAAGCATTCCAGTAG 1703
```

RESULT 14

AK026936

LOCUS

DEFINITION Homo sapiens cDNA: FLJ23283 fis, clone HEP08729, highly similar to AF006005 Homo sapiens zinc finger protein PLAGL2 mRNA.

ACCESSION

AK026936

VERSION

AK026936.1

GI:10439912

oligo capping; fis (full insert sequence).

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (sites)

Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,

Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,

Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 2429)

Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,

Shibahara, T., Tanaka, T. and Nakamura, Y.

Direct Submission

Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,

University of Tokyo, Laboratory of Genome Structure Analysis, Human

Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,

Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,

Fax:81-3-5449-5416)

NEDO human cDNA sequencing project supported by Ministry of

International Trade and Industry of Japan; cDNA full insert

sequencing. Research Association for Biotechnology; cDNA library

construction, 5' - & 3' -end one pass sequencing; Department of

Virology and Human Genome Center, Institute of Medical Science,

University of Tokyo (partly supported by Science and Technology

Agency).

Location/Qualifiers

1..2429

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="HEP08729"

/cell_line="HepG2"

/cell_type="hepatoma"

FEATURES

source

```
misc_feature 540 a 743 c 599 g 547 t
BASE COUNT 540 a 743 c 599 g 547 t
ORIGIN
Query Match 5.4%; Score 392.8; DB 9; Length 2429;
Best Local Similarity 58.3%; Pred. No. 7.4e-60;
Matches 826; Conservative 0; Mismatches 542; Indels 48; Gaps 6;
Qy 577 AACCTTCTCTGCAAGCTCTGACAGGCTTTAAAGTGTGAGAAATTAAGGTTTCAAC 636
Db 199 AAGTGCCAAATGTGAATTTTCGGGAACACCTTCTCAAAATGGGAGAAAGCTGAGGCTCAC 258
Qy 637 TCCTACTCTCACAGGAGAGAGGCTTACAAAGTGCATACAAAGAGCTGCACCAAGGCC 696
Db 259 AGCTCCCGCAACAGAGAGAGGACCATATAGTGCCTCAGCTGACACTGTGGCAAGGCT 318
Qy 697 TTTGTTTCTAAGTACAAATTAAGAGGACATGGTACTCATTTCTCTGAGAAACCCAC 756
Db 319 TTTGCTTCCAAATACAGCTGTATAGGCACATGCCACCATCAGCCAGCAAAACCCAC 378
Qy 757 AAGTGAATTTATGTGAGAAATGTTTACCCGAAAGATCATCTGAAGATCATCTCCAT 816
Db 379 CAGTGATGTGTGTATAGATGTTTACCCGAAAGACCATCTGGGAACCATCTGCAG 438
Qy 817 ACACAGACCTTAAACAAAGAGACGTTTAAAGTGGGAAGATGTGCGAAGACTTACAATACC 876
Db 439 ACCATGATCTTAAACAAAGAGGCTTCCACTGCTGTAGTGGTGAAGATTAACAATACG 498
Qy 877 AAGCTTGGATTTAAAGCTCACTTGGCTTGCATGCGCAACAAAGTGGTGCACCTCACTGT 936
Db 499 AAGCTGGGTACCGGCGCCACCTGGCCATGCTGCTGCCAGAGCGGTGCACCTCAGCTGC 558
Qy 937 AAGGTATGTTTGCMAACTTTTGAAGACGCGAGTGCTTCTGGAGCACCTTAAATCTCAT 996
Db 559 AAGGTGTGCTGCAGACCTTTGAGAGTACCCAGGCGCTGTAGAGCACCTGAAGGCCAC 618
Qy 997 GCAGGCAAGTCTCTGTGGGTTTAAAGAAAAAGACCACTGCGCAACATTTGTGATCGC 1056
Db 619 TCACGCGGTAGCAGGCGTCCAAAGAGAGAGACCCCTGTGACCACTGCGACCG 678
Qy 1057 CGTTTACACCCGAAAGAGTCCGAGACACATGTTGGTGCACACTGGAAGAAAGAC 1116
Db 679 CGTTTCTATCTCGTAAGGATGTACGCGGCACCTAGTGTGTCACACAGCGCGTAAGGAC 738
Qy 1117 TTCCTCTGCTAGTATTGTGCACAGATTTGGCGGAAAGGATCACTGACTCGACATATG 1176
Db 739 TTCCTGTGCCAGTACTGTGCCCTAGCGGTTTGGCCGTAAGGACCACTGACGGCTCATGTC 798
Qy 1177 AAGAGAGTCAAAATCAAGAGCTTCTGAAGTCAAAACAGAACAGCTGGATTTCCCTGAC 1236
Db 799 AAGAGAGCCACTCGCAGAGCTGCTCAAGATCAAGACAGAGCGCGTGGACATGTTAGCG 858
Qy 1237 CCATTACTGGAATGTGTGTGCTATAAAGACGAGCTCTTCCCGGTGATGCTCTTA 1296
Db 859 CTACTCAGCTGAGCTCCACAGTCAAGTGAAGGAGAGCTGAGCCCTGTGCTGTCATG 918
Qy 1297 CTTTCCAGTGA---ACTGTTTCAAGCCATTCAAAACACTTTGACGTTAAACCTCTAC 1353
Db 919 GCCTCTCGGAGCGTAATGGGACCAAGGCCCTTCCCTGGCATGTTGCCCATGGCATGAT 978
Qy 1354 AACACTCCATTTCAAGTCCATGACAGCTCGGATCTGCCCA---CCAATGATCAACT 1410
Db 979 GGTGCCACATCCCTACCATGCCAGCAGCGGCGTGCACACTCCCTGGTGCACAAACG 1038
Qy 1411 TTACTTTTGGGAATGACATGCCCAATAGATATGACACTGTTTCACTCCCTCTCACACCTT 1470
Db 1039 CTGCCCATGGTATGAGTACCTCCCTGGAATCCTCACCCTATCTTCTCCCGAGCTCAGCTC 1098
Qy 1471 TCITTTCAATATCGGTTCAAGTTCTACCTCATATGCAATTTCTATTCTCTGAAAAAGACAG 1530
```

```

Db      1099  CCTCAAATACAGCTGGATCTACCTCATCTTG-----CCGACAAA 1143
QY      1531  CCATTAAAGGGGAAATGAGAGTTTACCTGATGGAGTTACAAGTGGCGCTCTTCA 1590
Db      1144  TTGCCAAGTGGAGTGGATAGTTTCTGGCGGAGCTTCTCGGAAGCCTGTCCCTCA 1203
QY      1591  TCCCAAGATTCAGAGATCGTCATCATCAAGCTAGGTTGGATCTCAGATTGGTCC 1650
Db      1204  TCCGCTGAACCCAGCCGCTCA-----CCTCAGCCGGCGCA 1242
QY      1651  CTAGATGATGGTGCAGGACCTCTCCCTATCCAAAGCTCTATCTCCATCAGTGACCC 1710
Db      1243  GTGGGCGCTCTCTAGTAGAAGCACTGCTGCGAAGCCCGCCAACTCTCTAGGCC 1302
QY      1711  CTAACACACACAGCAATGGATTTTCTCAGTTGTTTAAATTTCAATCTTAAATGTCCT 1770
Db      1303  CTCTGGCTGCTAAATGTGACTTCTCCACCTACTGGGCTTCTTCCACTCAACTGCC 1362
QY      1771  CCTATAA---TCCTCTATCAGTGGGAGCTTGGATGAGCTATTCCAGGAAGCA 1827
Db      1363  CCGTGTAAACCCACTGGGCGCACAGGAGGCTGTCTATGGGCTACTCCAGGCTGAGGCA 1422
QY      1828  CATCTCTCTGTTCCAGCTCCCAACACACAGATCTTCCAGGATCTCTGCAAACT 1887
Db      1423  CAGCCCTGCTTACCACTTTGCAAGCTCAGCTCAGATTCCTCCAGAGCTGGGAGCA 1482
QY      1888  ATAGGGCTTGGGCTCTGCACTCAGCTGTCAGAGCTTTCCAGGAGTAAAGCAAGT 1947
Db      1483  CTGAACTTTGGGCTCTGCACTCTTGCTCTCTGTTTCAAGTCTGCTGAGT---AGC 1539
QY      1948  ACCACCTCCAGCTTCCATCAAGCTTTTCAGTAG 1983
Db      1540  ACCACCTGCTGCTTCCATCAAGCATTCAGTAG 1575

RESULT 15
LOCUS   AF006005 4051 bp mRNA linear PRI 07-MAR-2000
DEFINITION Homo sapiens zinc finger protein PLAGL2 (PLAGL2) mRNA, complete
cds.
ACCESSION AF006005
VERSION AF006005.1 GI:3513455
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4051)
AUTHORS Kas,K., Voz,M.L., Hensen,K., Meyen,E. and Van de Ven,W.J.
TITLE Transcriptional activation capacity of the novel PLAG family of
zinc finger proteins
JOURNAL J. Biol. Chem. 273 (36), 23026-23032 (1998)
MEDLINE 98389728
PubMed 9722527
REFERENCE 2 (bases 1 to 4051)
AUTHORS Kas,K., Hensen,K., Meyen,E., Voz,M.L. and Van de Ven,W.J.M.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-1997) Laboratory for Molecular Oncology, Center
of Human Genetics, Herestraat 49, Leuven 3000, Belgium
FEATURES
source
1..4051
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q11"
1..4051
/gene="PLAGL2"
203..1693
/gene="PLAGL2"
/note="C2H2-type zinc finger protein; PLAG-like 2"
/codon_start=1

gene
CDS
1099 1143
1531 1590
1144 1203
1591 1650
1204 1242
1651 1710
1243 1302
1711 1770
1303 1362
1771 1827
1363 1422
1828 1887
1423 1482
1888 1947
1483 1539
1948 1983
1540 1575


```

```

/product="zinc finger protein PLAGL2"
/protein_id="AAC34252.1"
/db_xref="GI:3513456"
/translation="MTTFTSVPMWIDAKOEVEVWKLVPKPRGREAESQVKQCEI
SGTPSPNGEKLPHSLPQEPQPSYCPOLHCGKAFASKVLYRHMAHSAOKPHOCMY
CDKPHRRKHLNHLQTHDPNKAULHSCCKKNYNTKLYRRHLAMHAAASGDLACKV
CLQTPFSTQALLHLKHAHSRRVAGAKKKPCDCHRRFYTRKRRHLVHTGRKD
FLCQYCAORFGRKHLRHVKSHQELLKIKTEFVMDLGLLSCSTVSVKLELSPVL
CMASRDVMTKAFPGMLPMWYGAHIPMSTPGVPHSLVHNTLPMGSPYLESSPIS
EAQLPKYQLGSTSYLPDKLPKVEYSLFLAELPGLSLSSAPQASPOPAAAAALLD
EAALLKSPANLSEALCAANVDFSHLLGELPLNLPNCNPGATGHLVMGYSQBAQPLL
TTLQAPQDPSGAPGALNFPGLHSLPFTSLSTTLPRHQAFQ"
BASE COUNT 874 a 1192 c 972 g 1013 t
ORIGIN
Query Match 5.4%; Score 392.8; DB 9; Length 4051;
Best Local Similarity 58.3%; Pred. No. 6.8e-60;
Matches 826; Conservative 0; Mismatches 542; Indels 48; Gaps 6;
QY 577 AACTTTCTTGGCCAACTGTGTGACAGGCTTTTAAACAGTGTGTGAGAAAATTAAAGTTTCAC 636
Db 317 AAGTGCATGTGAATTTTCGGGAACACCTTTCTCAATGGGAGAGAGCTGAGGCCTCAC 376
QY 637 TCTTACTCTCACACAGAGAGAGGCGCTTCAAGTGCATACAAAGACTGCACCAAGGCC 696
Db 377 AGCCTCCGCAACACAGAGAGAGACCATATAGCTGCCCTCAGCTGCACCTGTGCAAGGCT 436
QY 697 TTTGTTTCTTAAGTACAAATTAAGAGGCACATGGCTACTCAATCTCTCTGAGAAAACCCAC 756
Db 437 TTTGCTTCCAAATACAAAGCTGTATAGGCACTGGCCACCCACTCAGCCCAAGAACCCAC 496
QY 757 AAGTGTAAATATTGTGAGAAAATGTTTACCGGAAAGATCATCTGAAGATACCTCCAT 816
Db 497 CAGTGTATGACTGTGTAAGATGTTTACCGCAAGGACCATCTGCGAACCCTCTCAG 556
QY 817 ACACAGACCTTAAACAAAGAGACGTTTAAGTGCAGAGATGTGGCAAGACTACATAAC 876
Db 557 ACCCATGATCTTAAACAAAGAGGCGCTTCACTGTCTGAGTGCCTAAGAAATTAACAATACG 616
QY 877 AAGCTTGGATTAAACGTCATTTGGCTTGCATCGCAACAAAGTGGTGAACCTCACCTGT 936
Db 617 AAGCTGGGCTACCGGGCGCCACCTGGCCATCATCTGTCAGCAGCGGTGACCTCAGCTGC 676
QY 937 AAGGTATGTTTGCAAACTTTTGAAGCAGCGGAGTGTCTTGGAGCACCCTTAAATCTCAT 996
Db 677 AAGTGTGCTGCAGACCTTTGAGAGTACCCAGGCGCTGCTAGAGACCTGAAGGCCAC 736
QY 997 GCAGGCAAGTCGTCTGTGGGGTTTAAAGAAAAAAGAACACAGTGCAGCACTTGTGATCGC 1056
Db 737 TCACGCGGGTAGCAGCGCGTGCACAGGAGAAAGCACCCTGTGACCACTTGGACCGG 796
QY 1057 CGGTTCTACACCCGAAAGGATGTCGGAGACACATGCTGGTGCACACTGGAAGAAAGAC 1116
Db 797 CGGTTCTATCTCGTAAGGATGTACGGCGGCACTAGTGGTGCACACAGGCCGTAAGGAC 856
QY 1117 TTCCTCTGTCTCAGTATTGTGCAAGAGATTGGGGGAAAGGATACCTGACTGCAATATG 1176
Db 857 TTCCTGTGCCAGTACTGTGCCACGCGTTTGGCGGTAAAGGACCACTGACGCGTCAATGTC 916
QY 1177 AAGAGAGTCTCAATCAAGACTTCTGAAGTCAAAACAGAACACAGTGGATTCTCCTTGAC 1236
Db 917 AAGAAGAGCCACTCGCAGGAGCTCTCAAGATCAAGACAGAGCCCGTGGACATGTTAGGC 976
QY 1237 CCATTTTACCTGCAATGTGTCTGTGCTTATAAAGACGAGCTCTTCCGGTGTATGCTCTTA 1296
Db 977 CTACTCAGCTGCAGCTCCACAGTCACTGTGAGGAGAGGCTGAGCCCTGTCTGTGCAATG 1036
QY 1297 CCTTCCAGTGA---ACTGTTATCAAGCCATTCTCAAAACACTTTGACGTTAAACCTCTAC 1353
Db 1037 GCCTCTCGGGACGTAATGGGGACCAAGGCCCTTCCCTGGCATGTTGGCCATGGCATGAT 1096
QY 1354 AACACTCATTTCAGTCCATCGAGCTCGGGATCTGCCCA---CCAAATGATCACAACT 1410

```

Db 1097 GGTGCCACATCCCTACCATGCCCAGCAGCGGCGTGCCACACTCCCTGGTGCAACACG 1156
Qy 1411 TTACCTTTGGGAATGACATGCCCAATAGATATGGACACTGTTTCATCCCTCTCACACCTT 1470
Db 1157 CTGCCATGGGATGAGCTACCCCTCTGGAACTCCTCACCTATCTCTTCCCCAGCTCAGCTC 1216
Qy 1471 TCTTTCAAAATATCCGTTTCAGTTTCTACCTCATATGCAATTTCTATTCTTGAAAAAGAACAG 1530
Db 1217 CCTCCAAAATACCAGCTGGATCTACCTCATACTTG-----CCCGACAAA 1261
Qy 1531 CCAITTAAGGGGGAATTAAGAGTTTACCTGATGGAGTTTACAAGGTGGGTGCCCTCTTCA 1590
Db 1262 TTGCCCAAGTGGAGTGGATAGTTTCTTGGCGGAGCTTCTGGAAGCTGTCTCTCTCA 1321
Qy 1591 TCCCAAGATTCTCAAGCATGTCATCATCTAAGCTAGGTTGGATCCTCAGATTGGGTCC 1650
Db 1322 TCCGCTGAACCCAGCCCGCCTCA-----CCTCAGCCGGCGCA 1360
Qy 1651 CTAGATGATGTGSCAGGAGACCTCTCCCTATCCAAAAGCTTATCTCCATCAGTGACCCC 1710
Db 1361 GCTGGGCGCCTCCTAGATGAAGCACTGCTTGCCAGAGCCCGCCAACTCTCTGAGGCC 1420
Qy 1711 CTAAACACACAGCAATTGGATTTTCTCAGTTGTTTAATTTCATACCTTTAAAGTCTCT 1770
Db 1421 CTCTGCGCTGCTAATGTGGACTTCTCCACCTACTTGGGCTTCTTCCACTCAACCTGCCC 1480
Qy 1771 CCCTATTA--TCCTCTATCAGTGGGAGCCTTGGATGAGCTATTCGCCAGGAAGACA 1827
Db 1481 CCGTGTAAACCCACCTGGGGCCACAGGAGCCTGGTCTATGGGCTACTCCGAGGCTGAGCA 1540
Qy 1828 CATTTCTCTGTTTCCAGCTCCCCACACAAACACAGGATCTTCAGGATCCTGCMAACACT 1887
Db 1541 CAGCCCTGCTTACCACTTTGCAAGCTCAGCCTCAAGATTCCCCAGGAGCTGGGGGACCA 1600
Qy 1888 ATAGGCTTGGGTCTCTGCACTCACTGTCAAGAGCTTTCAAGCAGTTTAAAGCACAAGT 1947
Db 1601 CTGAACCTTGGGCGCTCTGCACTCTCTTGGCTCTCTGTCTTCAAGTCTGGGCTGAGT---AGC 1657
Qy 1948 ACCACCTCCCAAGTTTCCATCAGCTTTTCAGTAG 1983
Db 1658 ACCACCTGCGCTGCTTCCATCAGCATTCAGTAG 1693

Search completed: January 31, 2004, 13:36:46
Job time : 25657 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 03:59:43 ; Search time 1682 seconds
(without alignments)
11736.617 Million cell updates/sec

Title: US-09-242-772-116
Perfect score: 7313
Sequence: 1 ggcagcgatcacatcaaat.....tatgaataaaatctgtgpc 7313

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N: Geneseq 13Jun03.*
1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7312.6	100.0	7313	19	AAV29268 Nucleotide sequenc
2	351.6	4.8	2334	19	AAV18481 Human zinc finger
3	351.6	4.8	2828	24	ABN86525 Nucleotide sequenc
4	350	4.8	2815	24	ABL61932 Colon adenocarcino
5	350	4.8	4632	24	ABN86524 Nucleotide sequenc
6	301.8	4.1	5028	24	ABN86508 Rat Lot-1 polypept
7	294.4	4.0	2790	19	AAV18480 BOPI cDNA. Mus sp
8	294.4	4.0	2975	24	ABN86526 Nucleotide sequenc

9	277.6	3.8	2561	19	AAV29269	Nucleotide sequenc
10	277.6	3.8	2561	24	ABT11028	Human breast cance
11	276	3.8	2738	20	AAZ33549	Human breast tumou
c 12	267	3.7	475	22	ABA52644	Human foetal liver
c 13	267	3.7	475	22	ABA22432	Probe #898 for gen
c 14	267	3.7	475	22	AAK00905	Human brain expres
c 15	267	3.7	475	22	AAK26360	Human bone marrow
c 16	267	3.7	475	22	AAI10993	Probe #926 for gen
c 17	267	3.7	475	22	AAI32253	Probe #939 used to
c 18	267	3.7	475	22	AAI00914	Probe #905 used to
c 19	267	3.7	475	23	ABS25951	Human liver single
c 20	267	3.7	475	24	ABS00948	Human genome-deriv
c 21	131.4	1.8	230	22	ABA65256	Human foetal liver
c 22	131.4	1.8	230	22	ABA32357	Probe #10823 for g
c 23	131.4	1.8	230	22	AAK13673	Human brain expres
c 24	131.4	1.8	230	22	AAK39414	Human bone marrow
c 25	131.4	1.8	230	22	AAI20227	Probe #10160 for g
c 26	131.4	1.8	230	22	AAI45427	Probe #14113 used
c 27	131.4	1.8	230	22	AAI05931	Probe #5922 used t
c 28	131.4	1.8	230	23	ABS39002	Human liver single
c 29	131.4	1.8	230	24	ABS13501	Human genome-deriv
30	126.8	1.7	324	24	ABL77615	Human ovarian canc
31	74.6	1.0	8056	25	ABZ10246	Haematopoietic cel
32	72.4	1.0	12237	24	ABL34358	Human immune syste
33	72	1.0	72	22	AA513082	DNA encoding zinc
34	72	1.0	72	25	ACC41633	Human zinc finger
c 35	65.8	0.9	2864	22	AAH14677	Haematopoietic cel
c 36	65.4	0.9	7348	22	AA546336	Human cDNA sequenc
c 37	65.4	0.9	7348	25	ABZ10100	Tumour suppressor
c 38	64.8	0.9	8056	25	ABZ10100	Haematopoietic cel
c 39	63.6	0.9	113515	24	ABL34175	Human immune syste
40	62.8	0.9	9504	24	ABK28407	DNA transcription
41	62.4	0.9	703	22	AAH16178	Human cDNA clone (
42	62.4	0.9	2729	22	AAH16178	Human cDNA sequenc
43	62.4	0.9	5893	24	ABL32859	Human immune syste
44	61.6	0.8	6301	24	ABN80100	Human chemically m
45	61.4	0.8	5544	24	ABL70478	Chemically treated

ALIGNMENTS

RESULT 1
AAV29268
ID AAV29268 standard; cDNA; 7313 BP.
XX
AC AAV29268;
XX
DT 21-AUG-1998 (first entry)
XX
DE Nucleotide sequence of human PLAG1.
XX
KW Human PLAG1 gene; PLAG1; tumorigenesis gene; T-gene; PLAG2; CTNNB1;
KW antibody; benign tumour; malignant tumour; leukaemia; lymphoma;
KW cancer; inhibition; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 481...1983
FT /*tag= a
FT /product= "PLAG1 protein"
FT /*tag= b
FT /transl_except= (pos: 1603..1605, aa: Gly)
FT /*tag= c
FT /transl_except= (pos: 1861..1863, aa: Gly)

97EP-0200130.

PR 22-AUG-1996; 96EP-0202339.
XX (KULE-) KU LEUVEN RES & DEV.
PA (UWGO-) UNIV GOETTERBORGS HOLDINGBOLAGET AB.
PA (LEUV-) LEUVEN RES & DEV.
PI Kas KP, Stenman KGD, Van De Ven WJM, Voz ML;
XX P-PSDB; AAM37948.
XX WPI; 1998-132252/13.
XX DR P-PSDB; AAM37948.
XX
XX New tumorigenesis T-genes and proteins - useful for, e.g. preparing
PT antibodies for clinically diagnosing cells having non-physiological
PT proliferative capacity such as lipoblastomas
XX
XX Claim 2; Fig 4; 71pp; English.
XX
XX This is the nucleotide sequence of the human PLAG1 gene. It is a
CC tumorigenesis gene (T-gene), which is isolated in the form of PLAG1,
CC PLAG2, and CTNNB1 genes. Their proteins can be used as a starting
CC point for preparing antibodies for clinically/medically diagnosing
CC cells having a non-physiological proliferative capacity as compared
CC to wild type cells, where the former cells are selected from both
CC benign and malignant tumours, as well as leukaemia and lymphomas.
CC Derivatives of the T-gene are also used in the diagnosis and
CC preparation of therapeutical compositions for the treatment of cancers,
CC such as nucleic acid derivatives, and antibodies. The T-gene may be
CC used as a starting point for designing suitable expression-modulating
CC compounds or techniques for the treatment of non-physiological
CC proliferation phenomena in humans or animals. Expression inhibitors
CC of the T-gene can be used in the treatment of diseases involving
CC benign or malignant tumours.
XX
XX Sequence 7313 BP; 2287 A; 1327 C; 1356 G; 2342 T; 1 other;
SQ

Query Match 100.0%; Score 7312.6; DB 19; Length 7313;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 7312; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGCAGCGCATACACTACAATGGCTGCTGGAAGAGCGCTAAGAAACAATTCAGGCC 60
DB 1 GGCAGCGCATACACTACAATGGCTGCTGGAAGAGCGCTAAGAAACAATTCAGGCC 60
QY 61 GCGCGTCCAGCCGGAATATGAGAAATAATTATAGAAATTCGCGGCGGTAGAG 120
DB 61 GCGCGTCCAGCCGGAATATGAGAAATAATTATAGAAATTCGCGGCGGTAGAG 120
QY 121 GCGCGGAGCGGCGGAGGAGGATGTTAAAGCCCGCGTGTGCTCTTGGTGTGCTT 180
DB 121 GCGCGGAGCGGCGGAGGAGGATGTTAAAGCCCGCGTGTGCTCTTGGTGTGCTT 180
QY 181 GCGCGTATTTGGCAACCAGAAATGCTTCAATCTGTGACGGTCTAATAAGTTGCTTG 240
DB 181 GCGCGTATTTGGCAACCAGAAATGCTTCAATCTGTGACGGTCTAATAAGTTGCTTG 240
QY 241 CTAGAGTTTGGAGCAGGCCCTCAGATTGGCCAAATCGGAGGATTCGATTCACACTCT 300
DB 241 CTAGAGTTTGGAGCAGGCCCTCAGATTGGCCAAATCGGAGGATTCGATTCACACTCT 300
QY 301 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAAGTCTGAGGCTTTTCCATCAGTAAT 360
DB 301 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAAGTCTGAGGCTTTTCCATCAGTAAT 360
QY 361 CAGTCCCTTTTGTCTTTTACGACCAATGAAATCTGGAAGCCACCTAAGACTATA 420
DB 361 CAGTCCCTTTTGTCTTTTACGACCAATGAAATCTGGAAGCCACCTAAGACTATA 420
QY 421 TCATTTAGTGGAGTGGGAGTCCAGTCTCCACAGAGCGCTGTTTAGCTGCG 480
DB 421 TCATTTAGTGGAGTGGGAGTCCAGTCTCCACAGAGCGCTGTTTAGCTGCG 480
QY 481 ATGGCCACTGTCAATTCCTGGTGAATTTGTGAGAGTAAAGATACCCAGAAAGTCCCTTCA 540
DB 481 ATGGCCACTGTCAATTCCTGGTGAATTTGTGAGAGTAAAGATACCCAGAAAGTCCCTTCA 540

DB 481 ATGGCCACTGTCAATTCCTGGTGAATTTGTGAGAGTAAAGATACCCAGAAAGTCCCTTCA 540
QY 541 GGGAAACCTAAGCGTGGTGAACCAACCAAGAAAAAACTTTCTTGGCAACTGTGTGAC 600
DB 541 GGGAAACCTAAGCGTGGTGAACCAACCAAGAAAAAACTTTCTTGGCAACTGTGTGAC 600
QY 601 AAGGCTTTTAAACAGTGTGAGAAAAATAAGGTTCACTCTACTCTCACACAGGAGAGG 660
DB 601 AAGGCTTTTAAACAGTGTGAGAAAAATAAGGTTCACTCTACTCTCACACAGGAGAGG 660
QY 661 CCCTACAGTGCATACACAGACTGCACCAAGGCTTTTGTCTTAAGTACAAAAATACAA 720
DB 661 CCCTACAGTGCATACACAGACTGCACCAAGGCTTTTGTCTTAAGTACAAAAATACAA 720
QY 721 AGGCACATGGCTACTCAATCTCTGAGAAAAACCCACAGTGAATTAATTTGTGAGAAATG 780
DB 721 AGGCACATGGCTACTCAATCTCTGAGAAAAACCCACAGTGAATTAATTTGTGAGAAATG 780
QY 781 TTTTACCGGAAAGATCACTGAAGAAATCACTCTCTATACACAGCCCTAACAAAGAGCG 840
DB 781 TTTTACCGGAAAGATCACTGAAGAAATCACTCTCTATACACAGCCCTAACAAAGAGCG 840
QY 841 TTTTAAAGTGCAGAAATGTGGCAAGAACTAACAATACCAAGCTTGGATTTAAACCTCACTTG 900
DB 841 TTTTAAAGTGCAGAAATGTGGCAAGAACTAACAATACCAAGCTTGGATTTAAACCTCACTTG 900
QY 901 GCCTTGCATGCGCAACAAGTGGTGCACCTCACTGTAAGGTATGTTTGAACCTTTTGA 960
DB 901 GCCTTGCATGCGCAACAAGTGGTGCACCTCACTGTAAGGTATGTTTGAACCTTTTGA 960
QY 961 AGCAGCGGAGTGTCTTGGAGCACTTAAATCTCATGAGGCAAGTCTGTGGTGGGTT 1020
DB 961 AGCAGCGGAGTGTCTTGGAGCACTTAAATCTCATGAGGCAAGTCTGTGGTGGGTT 1020
QY 1021 AAGAAAAAAGCAGGTCGCAACTGTCGCGGTTCTTACACCCGAAAGAGTGC 1080
DB 1021 AAGAAAAAAGCAGGTCGCAACTGTCGCGGTTCTTACACCCGAAAGAGTGC 1080
QY 1081 CGGAGACATGCTGTGACACTGGAAGAAAGGACTTCTCTGTGATTTGTGCACAG 1140
DB 1081 CGGAGACATGCTGTGACACTGGAAGAAAGGACTTCTCTGTGATTTGTGCACAG 1140
QY 1141 AGATTTTGGCGAAAGGATCACTGACATGAGAAAGAGTCAATCAAGAGCTT 1200
DB 1141 AGATTTTGGCGAAAGGATCACTGACATGAGAAAGAGTCAATCAAGAGCTT 1200
QY 1201 CTGAGGTCGCAAGACAGGTCGATTCCTTGGACCATTTACCTGCAATGTGTGTG 1260
DB 1201 CTGAGGTCGCAAGACAGGTCGATTCCTTGGACCATTTACCTGCAATGTGTGTG 1260
QY 1261 CCTATAAAGACGAGTCTCTTCCGCTGATGTCCTTACCTTCCAGTGAACTGTTATCAAAG 1320
DB 1261 CCTATAAAGACGAGTCTCTTCCGCTGATGTCCTTACCTTCCAGTGAACTGTTATCAAAG 1320
QY 1321 CCATTCACAAACACTTTGCGATTAACCTCTAAGACTCCATTCAGTCCATGACAGC 1380
DB 1321 CCATTCACAAACACTTTGCGATTAACCTCTAAGACTCCATTCAGTCCATGACAGC 1380
QY 1381 TCGGATCTCGCCACCAAAATGATCACAATTTTGGAAATGACATGCCCAATAGAT 1440
DB 1381 TCGGATCTCGCCACCAAAATGATCACAATTTTGGAAATGACATGCCCAATAGAT 1440
QY 1441 ATGGACACTGTTCATCTCTCTCACCACTTTCTTCAAATATCCGTTCACTCTCA 1500
DB 1441 ATGGACACTGTTCATCTCTCTCACCACTTTCTTCAAATATCCGTTCACTCTCA 1500
QY 1501 TATGCAATTTCTATTCTCTGAAAAAGACAGCATTAAAGGGGAAATTTAGAGTTTACCTG 1560
DB 1501 TATGCAATTTCTATTCTCTGAAAAAGACAGCATTAAAGGGGAAATTTAGAGTTTACCTG 1560
QY 1561 ATGGAGTTACAGGTGGCGCTCTTTCATCCCAAGATTTCTCAAGATTCGTTCATCT 1620
DB 1561 ATGGAGTTACAGGTGGCGCTCTTTCATCCCAAGATTTCTCAAGATTCGTTCATCT 1620

1621 AAGCTAGGGTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGGAGACCTCTCCCTA 1680
1621 AAGCTAGGGTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGGAGACCTCTCCCTA 1680
1681 TCCAAAGAGCTCTATCTCCATCAGTGCAGCCCTTAAACACACAGAGATTGGATTTCCTCAG 1740
1681 TCCAAAGAGCTCTATCTCCATCAGTGCAGCCCTTAAACACACAGAGATTGGATTTCCTCAG 1740
1741 TTGTTTAAATTCATACCTTTAAATGGTCCCTCTAATAATCCTCTATCAGTGGGAGCCCTT 1800
1741 TTGTTTAAATTCATACCTTTAAATGGTCCCTCTAATAATCCTCTATCAGTGGGAGCCCTT 1800
1801 GGAATGAGCTATTCACAGGAGAGACCAATTCCTCTGTTTCCAGAGTCCCCACACAAACA 1860
1801 GGAATGAGCTATTCACAGGAGAGACCAATTCCTCTGTTTCCAGAGTCCCCACACAAACA 1860
1861 CAGGATCTTCAGGATCCTGCAACACTATAGGGCTTGGGTCTCTGCACTCACTGTGACGA 1920
1861 CAGGATCTTCAGGATCCTGCAACACTATAGGGCTTGGGTCTCTGCACTCACTGTGACGA 1920
1921 GCTTTCCAGCAGGATTTAAGCAACAAGTACCACTCCAGCTTCCAGCTCCCAAGCTTTTCAG 1980
1921 GCTTTCCAGCAGGATTTAAGCAACAAGTACCACTCCAGCTTCCAGCTCCCAAGCTTTTCAG 1980
1981 TAGGATCTGGGACATGGGATTCATTACAGAAATGTATGTAGTGTGCTGCTAGATGACC 2040
1981 TAGGATCTGGGACATGGGATTCATTACAGAAATGTATGTAGTGTGCTGCTAGATGACC 2040
2041 ATTATTTATTTAGTGGCTACTTTTAAACAGTATATAAAATTTCTGCTTTTGTATATACAA 2100
2041 ATTATTTATTTAGTGGCTACTTTTAAACAGTATATAAAATTTCTGCTTTTGTATATACAA 2100
2101 ATTTTCATTAAAGCCAGTATATAAATAGAACTAGCTTTTAAATGGAGCTTTGGAAACAATTT 2160
2101 ATTTTCATTAAAGCCAGTATATAAATAGAACTAGCTTTTAAATGGAGCTTTGGAAACAATTT 2160
2161 GTGTTTCAGTTAAGTTTACTGGGTATTTTGTCTGATTCAGTCCCAATTTGTACATTTTA 2220
2161 GTGTTTCAGTTAAGTTTACTGGGTATTTTGTCTGATTCAGTCCCAATTTGTACATTTTA 2220
2221 AGACTTTTATTTTCCATATAGGAAGCAATATTAGTAGTAAATCTTTTACAAATCCCA 2280
2221 AGACTTTTATTTTCCATATAGGAAGCAATATTAGTAGTAAATCTTTTACAAATCCCA 2280
2281 TTTTCAAAATTAATTTAGATCTTAAATTTTCAATTTTGTCTGATTCAGTCCCAATTTT 2340
2281 TTTTCAAAATTAATTTAGATCTTAAATTTTCAATTTTGTCTGATTCAGTCCCAATTTT 2340
2341 TTTTGACATCTGGCTCATTTAAAAATTTAGCAATAGAAATGTAAATTTGTATAAAAAGTTTG 2400
2341 TTTTGACATCTGGCTCATTTAAAAATTTAGCAATAGAAATGTAAATTTGTATAAAAAGTTTG 2400
2401 TGAATAACTCAAGGTTTAAATTTTCTTACTAGCTTCTAAATGGAATTAATCAAGTGC 2460
2401 TGAATAACTCAAGGTTTAAATTTTCTTACTAGCTTCTAAATGGAATTAATCAAGTGC 2460
2461 TTTCAAAATTAATTTAGATCTTAAATTTTCAATTTTGTCTGATTCAGTCCCAATTTT 2520
2461 TTTCAAAATTAATTTAGATCTTAAATTTTCAATTTTGTCTGATTCAGTCCCAATTTT 2520
2521 TTTTGACATCTGGCTCATTTAAAAATTTAGCAATAGAAATGTAAATTTGTATAAAAAGTTTG 2580
2521 TTTTGACATCTGGCTCATTTAAAAATTTAGCAATAGAAATGTAAATTTGTATAAAAAGTTTG 2580
2581 TTTATGAGCTTTAGGAAACCAAAACCTTAAACACAAAGCATCAGTATTTATAGCAAAAG 2640
2581 TTTATGAGCTTTAGGAAACCAAAACCTTAAACACAAAGCATCAGTATTTATAGCAAAAG 2640
2641 AGACTCTGTTAGGTGACATGGCATTTTGTGTCACCTTAATAGTTGGCCCTTAAATTTAGTACA 2700
2641 AGACTCTGTTAGGTGACATGGCATTTTGTGTCACCTTAATAGTTGGCCCTTAAATTTAGTACA 2700

2701 CAGGATATTTTGTGCTGTTTCACTCTTCTTAAACATGCTATCTTTTCAATTTAATAATAGTA 2760
2701 CAGGATATTTTGTGCTGTTTCACTCTTCTTAAACATGCTATCTTTTCAATTTAATAATAGTA 2760
2761 ATAGTGTATGGCATTTGGGGTCTTTCAGAGTCGATATATAGTAGATCTCTTTAGTCTTTTC 2820
2761 ATAGTGTATGGCATTTGGGGTCTTTCAGAGTCGATATATAGTAGATCTCTTTAGTCTTTTC 2820
2821 CACCTTTTCAATCAACAGGGGTGGGTCAAGTCGAGCCAGCAATTTATTTTCAATTTGTGGCC 2880
2821 CACCTTTTCAATCAACAGGGGTGGGTCAAGTCGAGCCAGCAATTTATTTTCAATTTGTGGCC 2880
2881 CACGGTGTAGTCCATATCTTAGGCCATTTGTGGAACCTGCGACCATAGGTGTGTATTTCCTC 2940
2881 CACGGTGTAGTCCATATCTTAGGCCATTTGTGGAACCTGCGACCATAGGTGTGTATTTCCTC 2940
2941 ACAGTGGATTGACTCAGCCTCTGTGGTGCACAGACTTCTTAAGCAGGAAGATAGACGTGAA 3000
2941 ACAGTGGATTGACTCAGCCTCTGTGGTGCACAGACTTCTTAAGCAGGAAGATAGACGTGAA 3000
3001 GCACATGGTTACATTTGGGAACCTTGTGTAGGGATCATGGCCCTGTAGCCAGGGTTAAAA 3060
3001 GCACATGGTTACATTTGGGAACCTTGTGTAGGGATCATGGCCCTGTAGCCAGGGTTAAAA 3060
3061 ACTGGACCTTTTAGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3120
3061 ACTGGACCTTTTAGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3120
3121 ATATGTTTTTCTTCCCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3180
3121 ATATGTTTTTCTTCCCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3180
3181 ATGTTATCTTCTGATTCATGTAATCAATCTTCTTCTGCAATATTTTGGATGAGAT 3240
3181 ATGTTATCTTCTGATTCATGTAATCAATCTTCTTCTGCAATATTTTGGATGAGAT 3240
3241 TTGAAGTTAAATGCAATAGAAAACCTA CTTGATGAACCTA CTTGATGAACCTA CTTGATGAACCTA 3300
3241 TTGAAGTTAAATGCAATAGAAAACCTA CTTGATGAACCTA CTTGATGAACCTA CTTGATGAACCTA 3300
3301 AAATATATACAGTAAATCCCACTTTCATGATCTCTGGGAAATGATAGGATATGCAA 3360
3301 AAATATATACAGTAAATCCCACTTTCATGATCTCTGGGAAATGATAGGATATGCAA 3360
3361 ATAAAGTTGAGTTTGTAGAGGGTAAACAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3420
3361 ATAAAGTTGAGTTTGTAGAGGGTAAACAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3420
3421 GAAATTAACAAATTCAGAAATATATATCTGAAATATATAGGCTTTTCTCCTACATTT 3480
3421 GAAATTAACAAATTCAGAAATATATATCTGAAATATATAGGCTTTTCTCCTACATTT 3480
3481 AGACAAACCAACATATCTTCTTAAAGCTTAAATTAATATATTTTCTTAAAGGTCAGTTGG 3540
3481 AGACAAACCAACATATCTTCTTAAAGCTTAAATTAATATATTTTCTTAAAGGTCAGTTGG 3540
3541 GACATTAACCTTAAAGAAACATATCTTATTAAGCAGTGTATTAACACCTTATTTTAAAGGTCAGTTGG 3600
3541 GACATTAACCTTAAAGAAACATATCTTATTAAGCAGTGTATTAACACCTTATTTTAAAGGTCAGTTGG 3600
3601 TCGTGTGGGATGGGGCAAGGGTGGAGGTTTTAGAGAGTATATATCTCTTAAAAA 3660
3601 TCGTGTGGGATGGGGCAAGGGTGGAGGTTTTAGAGAGTATATATCTCTTAAAAA 3660
3661 AAAACAGAAAGAAATATTTCTGAGCAGTCTATAGCCCTATATAGGAACTCTTTCTCTT 3720
3661 AAAACAGAAAGAAATATTTCTGAGCAGTCTATAGCCCTATATAGGAACTCTTTCTCTT 3720
3721 TTTGTAGGGCCAGTCTATCTGAGATTTGCAATGTTTACCAAGATTTCTTAAAAATGAGT 3780
3721 TTTGTAGGGCCAGTCTATCTGAGATTTGCAATGTTTACCAAGATTTCTTAAAAATGAGT 3780
3781 GCAGATTACTGAATATAATACATTTTAAAAATTTTGGGAGTAGTATAATTTGTTGAGA 3840

QY 6001 TTAGCAATGGCTGGCTGTTAGTTCTGGTAAATGTGCGCTAAGTTGAATTTGTTCTGTTT 6060
Db |||||
QY 6001 TTAGCAATGGCTGGCTGTTAGTTCTGGTAAATGTGCGCTAAGTTGAATTTGTTCTGTTT 6060
Db |||||
QY 6061 TTCTCACACTGTGTGAGCAGCCATGTCTCAACACAGATAAGTGTGTGATCACATAG 6120
Db |||||
QY 6061 TTCTCACACTGTGTGAGCAGCCATGTCTCAACACAGATAAGTGTGTGATCACATAG 6120
Db |||||
QY 6121 ATCTACATAAGTTGTGTCAGTTTGTGCTAAACCCCATAGGAGCTCTTTGGGATCATTA 6180
Db |||||
QY 6121 ATCTACATAAGTTGTGTCAGTTTGTGCTAAACCCCATAGGAGCTCTTTGGGATCATTA 6180
Db |||||
QY 6181 GAAAAAGAGATCATGCAACAGCAATTTGGTGAAGGCGCACCTCAGATTCAGTTAGGCGCTT 6240
Db |||||
QY 6181 GAAAAAGAGATCATGCAACAGCAATTTGGTGAAGGCGCACCTCAGATTCAGTTAGGCGCTT 6240
Db |||||
QY 6241 TCTATCATGTGTGTCACACCTCTGAGATGGAAGGCGAGTGTCTTTGATGTTATCTAGCCT 6300
Db |||||
QY 6241 TCTATCATGTGTGTCACACCTCTGAGATGGAAGGCGAGTGTCTTTGATGTTATCTAGCCT 6300
Db |||||
QY 6301 AGAAATGACACAGAACTATTGCTTAATGATATAAACAACCTTCATATATAAGCTTCAGTGGT 6360
Db |||||
QY 6301 AGAAATGACACAGAACTATTGCTTAATGATATAAACAACCTTCATATATAAGCTTCAGTGGT 6360
Db |||||
QY 6361 ACAGATGAACAGAAATGTAATGTTTATCTCTCAGAAACACTCTCTCAATATATATTGGA 6420
Db |||||
QY 6361 ACAGATGAACAGAAATGTAATGTTTATCTCTCAGAAACACTCTCTCAATATATATTGGA 6420
Db |||||
QY 6421 TCATGCTGTAATGTAACCTTGGGCTGACACTCTTCATGCTGTCACAACTCTCTGCTCTC 6480
Db |||||
QY 6421 TCATGCTGTAATGTAACCTTGGGCTGACACTCTTCATGCTGTCACAACTCTCTGCTCTC 6480
Db |||||
QY 6481 ATTCAGTCTGTAATTTTTTATCCATGAAAGAGACTACATTAGTGTAAAGTGTACAT 6540
Db |||||
QY 6481 ATTCAGTCTGTAATTTTTTATCCATGAAAGAGACTACATTAGTGTAAAGTGTACAT 6540
Db |||||
QY 6541 ATATTTTATAGTGTGACTTAATTTGTCATTAACAACTTTTACACCAACCAATGTAAT 6600
Db |||||
QY 6541 ATATTTTATAGTGTGACTTAATTTGTCATTAACAACTTTTACACCAACCAATGTAAT 6600
Db |||||
QY 6601 CATGTGCACTGCAAAAGAGATCTCGGACATGCAAAATGTTTACCAAGAACCAACCCAGCTT 6660
Db |||||
QY 6601 CATGTGCACTGCAAAAGAGATCTCGGACATGCAAAATGTTTACCAAGAACCAACCCAGCTT 6660
Db |||||
QY 6661 TTGTCCACAAAGTGTGACTGTAAGTGTGAAAGTGGGCTTTTATAAGGTGTGGAGT 6720
Db |||||
QY 6661 TTGTCCACAAAGTGTGACTGTAAGTGTGAAAGTGGGCTTTTATAAGGTGTGGAGT 6720
Db |||||
QY 6721 GAAGAACATGCTGTATGTTACTTAACAGCCCTTTGAAATTTTAAACAACTGGGAATCCATT 6780
Db |||||
QY 6721 GAAGAACATGCTGTATGTTACTTAACAGCCCTTTGAAATTTTAAACAACTGGGAATCCATT 6780
Db |||||
QY 6781 AGGAAACGGATGTCATCATCTGAACTGGAAGTGGGCTGCTGAAATGTAATTTTAGCT 6840
Db |||||
QY 6781 AGGAAACGGATGTCATCATCTGAACTGGAAGTGGGCTGCTGAAATGTAATTTTAGCT 6840
Db |||||
QY 6841 AATGAAAAAGTGTGTGACTAGTACTTAAAGTGTCTAATGATAAGTTTTCAGTCAA 6900
Db |||||
QY 6841 AATGAAAAAGTGTGTGACTAGTACTTAAAGTGTCTAATGATAAGTTTTCAGTCAA 6900
Db |||||
QY 6901 AATGAAAAAGAAAAAATCTGCAATTCAGGCGCAATTTGATATATTTTATTGCAATTTAA 6960
Db |||||
QY 6901 AATGAAAAAGAAAAAATCTGCAATTCAGGCGCAATTTGATATATTTTATTGCAATTTAA 6960
Db |||||
QY 6961 AATTTCTATTCTGTAATATTGGGAAATCAAGTGGCTTATCATGATATATCGTGTACTTAAA 7020
Db |||||
QY 6961 AATTTCTATTCTGTAATATTGGGAAATCAAGTGGCTTATCATGATATATCGTGTACTTAAA 7020
Db |||||
QY 7021 ATGTAATTCACAACTACTGTTGTAATTTGTAATATATATAGACAAAGATCATATTTTTTGT 7080
Db |||||
QY 7021 ATGTAATTCACAACTACTGTTGTAATTTGTAATATATATAGACAAAGATCATATTTTTTGT 7080
Db |||||

QY 7081 GTGTGATAAGCTCTGTAAATAGCAATCACTATGAAAGTGCAGTGATACTACATTTT 7140
Db |||||
QY 7081 GTGTGATAAGCTCTGTAAATAGCAATCACTATGAAAGTGCAGTGATACTACATTTT 7140
Db |||||
QY 7141 AAACATTACATCCAAAGAGCAGACTATTATTGTCATATACCAAGATTAAATATTA 7200
Db |||||
QY 7141 AAACATTACATCCAAAGAGCAGACTATTATTGTCATATACCAAGATTAAATATTA 7200
Db |||||
QY 7201 ATTTGCTGCTAAATTAATTAATAGTACTGCAAGCTTCTTGTGGCCTACAGTGTATGTTTC 7260
Db |||||
QY 7201 ATTTGCTGCTAAATTAATTAATAGTACTGCAAGCTTCTTGTGGCCTACAGTGTATGTTTC 7260
Db |||||
QY 7261 TGTAAAGATAAGTATGTGAATTTCCACAAATATATGAATAAATCTCGTGCC 7313
Db |||||
QY 7261 TGTAAAGATAAGTATGTGAATTTCCACAAATATATGAATAAATCTCGTGCC 7313
Db |||||

RESULT 2
AAV18481
ID AAV18481 standard; cDNA; 2334 BP.
XX
AC AAV18481;
XX
DT 18-AUG-1998 (first entry)
XX
DE Human zinc finger protein (hZAC).
XX
KW Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;
KW inducible; Alzheimer's disease; nuclear transcription factor; apoptosis;
KW cell cycle; neuronal disorder; human zinc finger; hZAC; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 803..2194
FT /*tag= a
FT /product= "hZAC"
XX
PN WO9813489-A1.
XX
PD 02-APR-1998.
XX
PF 22-SEP-1997; 97WO-EP05198.
XX
PR 23-SEP-1996; 96US-0718661.
XX
PA (CNRS) CENT NAT RECH SCI.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Journot L, Spengler D;
XX
DR WPI; 1998-230701/20.
XX
DR P-PSDB; AAW48761.
XX
PT New isolated tumour suppressor gene - useful for developing products
PT for use in diagnosis and treatment of tumour(s) or neuronal
PT disorder(s)
XX
PS Claim 17; Pages 82-85; 118pp; English.
XX
CC The present sequence represents the human zinc finger protein
CC (hZAC) cDNA isolated from the human pituitary gland cDNA library
CC using BOP1 cDNA (AAV18480) as a probe. Therefore the hZAC protein
CC is the human homologue of BOP1 (AAW48761). The hZAC protein displays
CC a tumour suppressing activity when it was constitutively and inducibly
CC expressed in tumour cells. The hZAC cDNA and the protein it encodes
CC are claimed to be useful in the preparation of therapeutic compositions,
CC useful for treating, preventing or delaying the recurrence of a tumour
CC or neuronal disorders, e.g. genetic diseases or acquired degenerative
CC encephalopathies such as Alzheimer's disease. The hZAC protein is
CC also claimed to be able to induce apoptosis resulting in inhibition
CC of tumour cell growth, to suppress tumour formation, to induce G1
CC arrest of the cell cycle and to act as nuclear transcription factor.

	RESULT	ABN86525	ABN	ABN	21-	Nuc	Cer cer	Hom	WO2
	ID	XX							
	AC	XX							
	DT	XX							
	XX								
	DE								
	XX								
	KW								
	KW								
	XX								
	O S								
	XX								
	P N								
	XX								

PT Lee RT;
 XX
 DR WPI; 2002-590446/63.
 XX
 XX Diagnosing cardiovascular conditions e.g. stroke, myocardial infarction
 PT as a subject, by detecting expression of a nucleic acid molecule such
 PT as Fit-1, CD44, Lot-1, AA892598 and Mrg-1 in biological sample from
 PT subject -
 XX
 PS Disclosure; Page 105-107; 113pp; English.
 XX
 CC The invention relates to diagnosing a cardiovascular condition
 CC characterised by aberrant expression of a nucleic acid molecule (I) such
 CC as Fit-1, CD44, Lot-1, AA892598 and Mrg-1 or its expression product. One
 CC method involves contacting a biological sample from a subject with an
 CC agent which specifically binds to (I), its expression product or a
 CC fragment of an expression product and measuring the amount of bound
 CC agent. The method is useful for diagnosing a cardiovascular condition
 CC such as myocardial infarction, stroke, arteriosclerosis, heart failure,
 CC and cardiac hypertrophy. Methods useful for determining the stage of
 CC cardiovascular condition and pharmaceutical compositions for treating the
 CC above cardiovascular conditions are also provided. The present sequence
 CC represents a human Lot-1 cDNA.
 XX
 SQ Sequence 4632 BP; 1232 A; 1168 C; 988 G; 1243 T; 1 other;

RESULT 6	
ABN85508	
ID	ABN86508 standard; cDNA; 5020 BP.
XX	
AC	ABN86508;
XX	
DT	21-OCT-2002 (first entry)
XX	
DE	Rat Lot-1 polypeptide encoding cDNA.
XX	
KW	Cardiovascular; Pit-1; CD44; Lot-1; AA892598; Mrg-1; cardiant; rat;
KW	cerebroprotective; antiarteriosclerotic; gene therapy; gene; ss.
XX	
OS	Rattus norvegicus.
XX	
FH	Key
FT	Location/Qualifiers
FT	548..2299
FT	/tag= a
FT	/product= "Lot-1"
XX	
PN	WO200238794-A2.
XX	

D_B 475 ACCTGGAAGAACCACTCTCAGACCCACCGATLCCACACAGAATCTCCATCCGGTGCACCATTA
QY 857 GTGCCAAGAACTACAATAACCAAGCTTGGAATTTAAAGTCATTGGCCCTTGGATCCGGCAA 916
D_B 535 GCGCCAAGAAGTACACACACCATGTGCGGCTACAGAGGCGCACTTGGCCCTTGCACTCGGCGA 594
QY 917 CAGTGTGTGACCTCACCTGTAAAGTAGTTGTTCAAAATTTTTGAAAGCACAGGAGTGTCTTC 976
D_B 595 GCATGGCGATCTCACCTGTGGGTGTGCACCTTGGAGCTGGGAGCACCGAGTCTCTCGC 654
QY 977 TGGAGCACCTTAATCTCATGC----AGGCAAGTCGTGTGGTGGGTTTAAAGAAAAAAGC 1033
D_B 655 TGACCACCTCAAGTCTCACGCGAAGAAAAGGCCAACGAGCACCCGAGGAAGAAT 714
QY 1034 ACCAGTGGCAACATTGTGATCGCGGTTTCTACACCGAAGAGATGTCCGAGACACATGG 1093
D_B 715 ACCAGTGGCGACCTGTGATAGATGCTTCTACACCGGAAGATGTGGCTGCCACCTGG 774
QY 1094 TGGTGCACACTGGAAAGAAGCACTTCCTCTGTAGTATTGTGCACAGAGATTTGGCGCAA 1155

D_B 475 ACCTGGAAGAACCACTTCACAGACCACCGATLCCACAAAGAATCTCCTACGGCTGCACAGTAA
QY 857 GTGCCAAGAACTACAATAACCAAGCTTGGAATTTAAAGCTCACTTGGCCCTTGGATGCCGAA 916
D_B 535 GCGCCAAGAAGTACACACCATGTGTGGGTACAAAGAGGCACCTGGCCCTGCACCTCGCGCA 594
QY 917 CAGTGTGTGACCTCACCTGTAAAGTAGTTGTTCAAAATTCTTTGAAAGCACAGGAGTGTCTTC 976
D_B 595 GCATGGCGATCTCACCTGTGGGTGTGCACCTTGGAGCTGGGAGCACCGAGTCTCTGC 654
QY 977 TGGAGCACCTTAATCTCATGC----AGCAGTCTGTGTGGGGTTAAAGAAAAAAGC 1033
D_B 655 TGACCACTCAAGTCTCACGGGAAGAAAGGCCAACGAGCACCCGAGGAAGAAT 714
QY 1034 ACCAGTGGCAACATTGTGATCGCGGTTTCTACACCGAAGAGATGTCCGAGACACATGG 1093
D_B 715 ACCAGTGGGACCACTGTGATAGTCTTCTACACCGGAAGATGTGGCTGCCACCTGG 774
QY 1094 TGGTCCACACTGGAAAGAAGCACTTCCTCTGTAGTATTGTGCACAGAGATTGGCGCAA 1155

Db 775 TGGTCCACAGGATGCAAGGACTTCTGTGTCTAGTCTGTGCCACAGAGATTGGGGCGCA 834
 QY 1154 AGGATCCTGACTCGACATATGAGAGAGTCAATCAAGAGCTTCTGAAGTCAAAA 1213
 Db 835 AAGACCACTCTCACTCGTCAACACCAAGAGACCCACTCCAGAGCTGATGCAAGAGAATA 894
 QY 1214 CAGAACCACTGATTTCTTTCGACCCATTACTCTGCAATGTCT 1257
 Db 895 TGCAGGCAGGAGATTACCAAGCAATTTCCAATCTATTGGCGCT 938

RESULT 9 AAV29269

ID AAV29269 standard; cDNA; 2561 BP.

AC AAV29269;

XX 21-AUG-1998 (first entry)

DE Nucleotide sequence of human PLAG2.

XX Human PLAG2 gene; PLAG2; tumorigenesis gene; T-gene; PLAG1; CTNNB1;
 KW antibody; benign tumour; malignant tumour; leukaemia; lymphoma;
 KW cancer; inhibition; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH 177..1416
 FT CDS /tag= a
 FT /product= "PLAG2 protein"

XX EP825198-A1.
 XX 25-FEB-1998.
 XX 17-JAN-1997; 97BP-0200130.
 XX 22-AUG-1996; 96BP-0202339.
 XX (KULE-) KU LEUVEN RES & DEV.
 PA (UYGO-) UNIV GOETTERBORGS HOLDINGBOLAGET AB.
 PA (LEUV-) LEUVEN RES & DEV.

PI Kas KP, Stenman KGD, Van De Ven WJM, Voz ML;
 DR WPI; 1998-132252/13.
 DR P-PSDB; AAW37949.
 XX New tumorigenesis T-genes and proteins - useful for, e.g. preparing
 PT antibodies for clinically diagnosing cells having non-physiological
 PT proliferative capacity such as lipoblastomas
 XX Claim 4; Fig 8; 71pp; English.

XX This is the nucleotide sequence of the human PLAG2 gene. It is a
 CC tumorigenesis gene (T-gene), which is isolated in the form of PLAG1,
 CC PLAG2, and CTNNB1 genes. Their proteins can be used as a starting
 CC point for preparing antibodies for clinically/medically diagnosing
 CC cells having a non-physiological proliferative capacity as compared
 CC to wild type cells, where the former cells are selected from both
 CC benign and malignant tumours, as well as leukaemia and lymphomas.
 CC Derivatives of the T-gene are also used in the diagnosis and
 CC preparation of therapeutic compositions for the treatment of cancers,
 CC such as nucleic acid derivatives, and antibodies. The T-gene may be
 CC used as a starting point for designing suitable expression-modulating
 CC compounds or techniques for the treatment of non-physiological
 CC proliferation phenomena in humans or animals. Expression inhibitors
 CC of the T-gene can be used in the treatment of diseases involving
 CC benign or malignant tumours.

XX Sequence 2561 BP; 704 A; 624 C; 516 G; 716 T; 1 other;

Query Match 3.8%; Score 277.6; DB 19; Length 2561;
 Best Local Similarity 71.9%; Pred. No. 6.2e-43;
 Matches 377; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

QY 721 AGGCACATGCTACTCTCTCTGAGAAAACCCACAGTGTATTTGTGAGAAAATG 780
 Db 171 AGGCATATGCTATCCCAATTTCTCCCAAGAAATCTCACCAGTGTGCTCACTGTGAGAGACG 230
 QY 781 TTTTACCGGAAAGATCATCTGAAGAATCACCTCATACACAGACCCCTACAAAGACG 840
 Db 231 TTCAACCGGAAAGACCACTGAAAACCACTCCAGACCACCCCAACCAATGGCC 290
 QY 841 TTTAAGTCCGAAGATGTGCAAGAACTACAATACCAAGCTTGGATTAAACGTCACCTG 900
 Db 291 TTTGGGTGTGAGGAGTGTGGGAAGAAGTACAACACCATGCTGGGCTATAAGAGCCACTG 350
 QY 901 GCCTTCATGCGCGCAACAGTGTGACCTCACTGTAAGTATGTTTGCAAACTTTTGA 960
 Db 351 GCCTTCATGCGCGCGAGCTGGGGACCTCACCTGTGGGTCTGTGCGCTGGAGCTAGGG 410
 QY 961 AGCAGCGGAGTGTCTCTGGAGCACCTTAAATCTCATGC--AGGCAAGTCTGTGTGGG 1017
 Db 411 AGCAGCGAGTGTCTACTTGGACCACTCAAGCCCATGCGGAAGAGAGGCCCTAGCGGA 470
 QY 1018 GTTAAAGAAAAAAGCACCAAGTCCGCAACTTGTGATCGCGGTCTTACACCCGAAAGGAT 1077
 Db 471 ACCAAGAAAAAAGAACCACTGCGACCACTGTGAAAGATGCTTCTACACCCGAAAGGAT 530
 QY 1078 GTCCGAGACACATGTTGTGTGCACACTGGAAGAAAGACTTCTCTGTGAGTATTTGCA 1137
 Db 531 GTGCGAGCCACTGTGTGTGTCACACAGGATGCAAGGACTTCTGTGCGAGTCTCTGTC 590
 QY 1138 CAGAGATTGGCGGAAAGGATCACCTGACTCCACATATGAAGAAGTGCACATCAAGAG 1197
 Db 591 CAGAGATTGGCGGCAAGGTTCACTCACCAGCATACCAAGAGACCCACTCAGAGAG 650
 QY 1198 CTTCTGAAGGTCAAAACAGAACCACTGGATTTCCTTGACCCCAAT 1241
 Db 651 CTGATGAAAGAGAGCTTGCAGACCGAGACCTTCTGAGCACCTT 694

RESULT 10

ABT11028

ID ABT11028 standard; cDNA; 2561 BP.

XX AC ABT11028;

XX XX

DT 04-DEC-2002 (first entry)

DE Human breast cancer associated coding sequence SEQ ID NO: 1162.

XX Human; breast specific gene; breast cancer; differential expression;
 KW cytostatic; gene therapy; gene; ss.

XX OS Homo sapiens.

XX XX WO200259271-A2.

XX XX 01-AUG-2002.

XX XX 25-JAN-2002; 2002WO-US02176.

XX XX 25-JAN-2001; 2001US-263757P.

XX XX 25-APR-2001; 2001US-286090P.

XX XX 23-MAY-2001; 2001US-292517P.

XX XX (GENE-) GENE LOGIC INC.

XX XX Orr MS, Nation M, Diggans JC, Zeng W;
 XX WPI; 2002-674803/72.

PT Diagnosing breast cancer in a patient comprises detecting the level of

PT gene expression in cell or tissue samples, where a differential gene
 PT expression is indicative of breast cancer -
 XX
 PS Claim 1; SEQ ID NO 1162; 260pp + Sequence Listing; English.
 XX
 CC The present invention relates to methods of diagnosing breast cancer in a
 CC patient, which comprise detecting the level of expression in a tissue
 CC sample of two or more genes selected from those shown in AB09867-
 CC AB01112, where a differential expression of the genes indicates breast
 CC cancer. The methods are useful in diagnosing, treating, detecting the
 CC progression, and in monitoring treatment of breast cancer in patients.
 CC The methods are also useful as a screening tool for agents that modulate
 CC the onset or progression of breast cancer. The breast cancer genes may be
 CC used as diagnostic markers for the prediction or identification of the
 CC malignant state of breast tissue, for confirming the type and progression
 CC of cancer, and for drug screening and assays. The present sequence is a
 CC coding sequence of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2561 BP; 704 A; 624 C; 516 G; 717 T; 0 other;

Query Match 3.8%; Score 277.6; DB 24; Length 2561;
 Best Local Similarity 71.9%; Pred. No. 6.2e-43;
 Matches 377; Conservative 0; Mismatches 144; Indels 3; Gaps 1;
 QY 721 AGGCATGCTACTCATCTCTGAGAAACCCACAGTGTAAATTTGAGAAATG 780
 Db |||||
 QY 171 AGGCATGCTACTCATCTCTGAGAAACCCACAGTGTAAATTTGAGAAATG 780
 Db |||||
 QY 781 TTTCACCGGAAAGACCACTTCCCGAATCTCCACAGTGTCTCACTGTGAGAGACG 230
 Db |||||
 QY 231 TTTCACCGGAAAGACCACTTCCCGAATCTCCACAGTGTCTCACTGTGAGAGACG 290
 Db |||||
 QY 841 TTTCAGTGCAGAGATGTCGAGAACTCAATATCAAGCTTGGATTTAAACGTCACCTG 900
 Db |||||
 QY 291 TTTCAGTGCAGAGATGTCGAGAACTCAATATCAAGCTTGGATTTAAACGTCACCTG 350
 Db |||||
 QY 901 GCCTTGATCGCGCAACAGTGTGACCTCACCTGTAGGATGATTTGGCAACTTTTGA 960
 Db |||||
 QY 351 GCCTTGATCGCGCAACAGTGTGACCTCACCTGTAGGATGATTTGGCAACTTTTGA 410
 Db |||||
 QY 961 AGCACGGAGTGTCTTGTGAGACCTTAAATCTCATGC---AGCAAGTCTCTGTGGG 1017
 Db |||||
 QY 411 AGCACGGAGTGTCTTGTGAGACCTTAAATCTCATGC---AGCAAGTCTCTGTGGG 470
 Db |||||
 QY 1018 GTTAAAGAAAGAACCCAGTGTGCAATGTCATGCGCGTTCTACACCCGAAAGGAT 1077
 Db |||||
 QY 471 ACCAAGGAAAGAACCCAGTGTGCAATGTCATGCGCGTTCTACACCCGAAAGGAT 530
 Db |||||
 QY 1078 GTCCGAGACATGCTGTGTCACATGTCGAGAAAGGACTTCTCTGTGATTTGTGCA 1137
 Db |||||
 QY 531 GTGAGCCACCTGTGTGTCCACACAGGATGCAAGGACTTCTCTGTGCGAGTCTGTGCC 590
 Db |||||
 QY 1138 CAGAGATTTGGCGGAAAGATCACTGATCTGATATGAGAGAGAGTCAATCAAGAG 1197
 Db |||||
 QY 591 CAGAGATTTGGCGGAAAGATCACTGATCTGATATGAGAGAGAGTCAATCAAGAG 650
 Db |||||
 QY 1198 CTTCTGAAGGTCAAAACAGAACCCAGTGTGATTTCTTGTGACCCATT 1241
 Db |||||
 QY 651 CTGATGAAGAGAGCTTGCAGACCGGAGACCTTCTGAGGACCTT 694
 Db |||||

RESULT 11
 AAZ33549
 ID AAZ33549 standard; cDNA; 2738 BP.
 XX
 AC AAZ33549;
 XX
 DT 08-DEC-1999 (first entry)
 XX
 DE Human breast tumour-associated EST 9.

XX
 KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;
 KW medicaments; gene therapy; treatment; fat metabolism; ss.
 XX
 OS Homo sapiens.
 XX
 PN DE19813835-A1.
 XX
 PD 23-SEP-1999.
 XX
 PF 20-MAR-1998; 98DE-1013835.
 XX
 PR 20-MAR-1998; 98DE-1013835.
 XX
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 PI Specht T, Hinzmann B, Schmitt A, Pilarzsky C, Dahl E, Rosenthal A;
 XX
 DR WPI; 1999-528979/45.
 DR P-PSDB; AAY48468.
 XX
 PT Human nucleic acid sequences and protein products from normal breast
 PT tissue, useful for breast cancer therapy -
 XX
 Claim 3; 96-97; 206pp; German.
 XX
 CC This invention describes novel human nucleic acid sequences from normal
 CC breast tissue which have cytostatic activity. The nucleic acid sequences
 CC can be used to produce and isolate full-length gene sequences. They can
 CC be used to express proteins, which can be used as tools to find an
 CC activity against breast cancer. The sequences can be used in sense or
 CC antisense form. They are especially useful for medicaments for gene
 CC therapy to treat breast cancer and for treating illnesses associated
 CC with fat metabolism. AA233541-233610 represent expressed sequence tags
 CC described in the method of the invention.
 XX
 SQ Sequence 2738 BP; 771 A; 661 C; 546 G; 760 T; 0 other;

Query Match 3.8%; Score 276; DB 20; Length 2738;
 Best Local Similarity 71.2%; Pred. No. 1.3e-42;
 Matches 379; Conservative 0; Mismatches 150; Indels 3; Gaps 1;
 QY 713 AATTACAAAGGCATCGCTACTCTCTCTGAGAAACCCACAGTGTAAATTTATGTG 772
 Db |||||
 QY 324 AATCAGTAGGCAATTTGGTACCCATTCTCCCCAGAAATCTCACAGTGTCTCACTGTG 383
 Db |||||
 QY 773 AGAAATGTTTCCCGGAAAGATCATCTGAAGATCACTCTCCATPACACAGCCTTAACA 832
 Db |||||
 QY 384 AGAAGAGCTTCAACCGGAAAGACCCCTGAAACACCTCCAGACCCACCAACA 443
 Db |||||
 QY 833 AAGAGAGCTTTAAGTGGCAAGATGTGGCAAGACTCAATPACCAAGCTTGGATTAAAC 892
 Db |||||
 QY 444 AATTTGGCTTTGGTGTGAGAGTGTGGAAAGAGTACACACCATCTGGGCTATAAGA 503
 Db |||||
 QY 893 GTCACTTGGCTTGGATGCCCAACAGTGTGTGACCTCACCTGTAAGGTATGTTTGCAAA 952
 Db |||||
 QY 504 GGCACCTGGCCCTCCATCGCCGACAGTGGGACCTCACCTGTGGGTCTGTGCCCTTGG 563
 Db |||||
 QY 953 CTTTGTGAAGCAGCGAGTGTCTTGGAGACCTTAAATCTCATGC---AGSCAAAGTCTG 1009
 Db |||||
 QY 564 AGCTAGGGAGCAGCCGAGTGTCTTGGACCTCAAGAGCCCATGCGAAGAGAGAGCC 623
 Db |||||
 QY 1010 CTGTTGGGTTTAAAGAAAAAGCAACAGTGTGCAATTTGATCGCGGTTCTACACC 1069
 Db |||||
 QY 624 CTAGCGGAACCAAGAAAGAGCAGCAGTGTGACCTGTGAAGATGCTTCTACACC 683
 Db |||||
 QY 1070 GAAAGAGTGTCCGAGACACATGTTGGTGTGACCTTGGAGAGAGGACTTCTCTGTCACT 1129
 Db |||||
 QY 684 GGAAGGATGTGCGACGCCACCTGTGGTCCACACAGGATGCAAGGACTTCTGTGCCACT 743
 Db |||||
 QY 1130 ATTGTGCAAGAGATTTGGCGGAAAGGATCACTGACTGCATATGAGAGAGAGTCA 1189
 Db |||||
 QY 744 TCTGTGCCAGAGATTTGGCGCAGAGGATCACTTCCACCGGCATACCAAGAGAGCCACT 803
 Db |||||

QY 1190 ATCAAGAGCTTCTGAAGGTCAAAACAGAACCAAGTGAATTCCTTGACCCATT 1241
 Db 804 CACAGGAGCTGATGAAGAGAGCTGCAGACCGGAGACCTTCTGAGCACCTT 855

RESULT 12
 ABA52644/c
 ID ABA52644 standard; DNA; 475 BP.
 XX ABA52644;
 AC ABA52644;
 DT 01-FEB-2002 (first entry)
 XX Human foetal liver single exon nucleic acid probe #949.
 DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX Homo sapiens.
 OS Homo sapiens.
 PN WO200157277-A2.
 XX 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00669.
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-483447/52.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX Claim 1; SEQ ID NO 949; 639pp + sequence listing; English.
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 475 BP; 75 A; 129 C; 142 G; 129 T; 0 other;
 SQ

Query Match
 Best Local Similarity 3.7%; Score 267; DB 22; Length 475;
 Matches 352; Conservative 74.1%; Pred. No. 4.3e-41;
 Mismatches 120; Indels 3; Gaps 1;

QY 733 ACTCATTTCTCTGAGAAACCCACAAAGTGTATTAATTGTGAGAAATGTTTCACCGAAA 792
 Db 475 ACCATTCTCCCGAAGAAATCCACAGTGTGCTCACTGTGAGAGACGTTCAACCGGAAA 416
 QY 793 GATCATCTGAAGAATCACTCCATACACACACACCTTAACAAAGAGACGTTTAACTGGCAA 852
 Db 415 GACCACCTGAANAACCTCCAGACCCACACCCCAACAAATGCGCTTTGGGTGTGAG 356
 QY 853 GAATGTGGCAGAACTCAATATACCAAGCTTGGATTAAACGTCACTTGGCCCTTCATGCC 912
 Db 355 GAGTGTGGGAGAGTACACACCATCTGGGCTTATAAGAGGCACCTGGCCCTCCATGCG 296

QY 913 GCAACAAGTGGTGACCTCACCTGTAAAGGTATGTTTGCAAACTTTTGAAAGCACGGAGNG 972
 Db 295 GCCAGCAGTGGGGACCTCACCTGTGGGTCTGTGCTCCCTGGAGCTAGGAGACCCGAGGTG 236
 QY 973 CTTCTGGAGCACCTTAATCTCATGC---AGCAAGTGTCTGTGGTGGGTTAAAGAAAAA 1029
 Db 235 CTACTGGACCACTCAAGCCCATGCGGAAGAGAACCCCTTAGCGAACCAAGGAAAG 176
 QY 1030 AAGCACCAAGTGCAGACATTGTGATCGCGGTCTACACCCGAAAGAGATGTCGGGAGACAC 1089
 Db 175 AAGCACCAAGTGCAGACCTGTGAAAGATGCTTACACCCGAAAGAGATGTCGGACGCCAC 116
 QY 1090 ATGTTGGTGACACTGGAAGAAAGACTTCTCTGTCACTATTTGTGCACAGAGATTTCGG 1149
 Db 115 CTGTTGGTCCACACAGGATGCAAGACTTCTGTGCCAGTTCTGTGCCACAGATTTCGG 56
 QY 1150 CGAAAGGATCATCGACTCGACATATGAAGAGAGTCAATCAAGAGCTTCTGA 1204
 Db 55 CGCAAGGATCACCTCACCCGCGCATACCAAGAGACCCCACTCACAGGAGCTGATGA 1

RESULT 13

ABA22432/c

ID ABA22432 standard; DNA; 475 BP.

XX ABA22432;

AC ABA22432;

DT 23-JAN-2002 (first entry)

DE Probe #898 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease; ss.

XX Homo sapiens.

OS Homo sapiens.

XX WO200157274-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -

XX Claim 1; SEQ ID No 898; 530pp; English.

The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO

at ftp.wipo.int/pub/published_pct_sequences.

Sequence 475 BP; 75 A; 129 C; 142 G; 129 T; 0 other;

Query Match 3.7%; Score 267; DB 22; Length 475;
Best Local Similarity 74.1%; Pred. No. 4.3e-41;
Matches 352; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

733 ACTCATCTCTCGAGAAACCCACCAAGTGTAAATTATCTGGAGAAATGTTTCACCGGAAA 792
475 ACCCATTTCTCCACAGAAATCTCACAGTGTGCTCACTGTGAGAAAGACGTTTCACCGGAAA 416

793 GATCATCTGGAAGAAATCACCTCCATACACACGACCCCTAACAAAGAGACGTTTAAAGTCGAAA 852
415 GACCACCTGAAAACCACTCCAGACCCACGACCCCAACAAATGGCCTTTGGGTGTGAG 356

853 GAATGTGGCAGAAGCTACAAATACCAAGCTTGGATTAAACGTCACCTGGCCCTTGCATGCC 912
355 GAGTGTGGGAAGAAGTACACACCACTGCTGGGCTATAAGAGGCACTTGGCCCTCCATGCG 296

913 GCACAAGTGGTGACCTCACCTGTAAGGTATGTTTGCAAACTTTTGAAAGACACGGGAGTG 972
295 GCACAGTGGGGACCTCACCTGTGGGTCTGTGCCCTGGAGCTAGGAGCACCGAGGTG 236

973 CTTCTGGAGCACCTTAAATCTCATGC---AGGCAAGTGTCTGTGGGTAAAGHAAA 1029
235 CTACTGGACCACCTCAAAGCCCATGGGAAGAACCCCTAGCGGAACCAAGGAAAAG 176

1030 AAGCACAGTGGCAACATTTGTGATCGCGGTCTTACACCGAAAGGATGTCGGGAGACAC 1089
175 AAGCACAGTGGCACCACTGTGAAGATGTTCTTACCCCGAAGATGTGCGAGCCAC 116

1090 ATGTGTGTGCACACTGGAAGAAAGGACTTCTCTGTCAAGTATTGTGCACAGATTTGGG 1149
115 CTGTGTGTGCACACAGATGCAAGGACTTCTCTGTGCCAGTTCTGTGCCAGAGATTTGGG 56

1150 CGAAGGATCACCTGATCTGACATATGAAGAAGATCACATCAAGAGCTTCTGA 1204
55 CGCAAGGATCACCTCACCGGCATACCAAGAAGACCCACTCACAGGAGTGTATGA 1

RESULT 14
AAK00905/c
ID AAK00905 standard; DNA; 475 BP.
XX AC AAK00905;
XX AC AAK00905;
XX AC AAK00905;
DT 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe SEQ ID NO: 896.
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PN 09-AUG-2001.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.

```
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
PS Example 4; SEQ ID NO: 917; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 475 BP; 75 A; 129 C; 142 G; 129 T; 0 other;
```

Query Match 3.7%; Score 267; DB 22; Length 475;
Best Local Similarity 74.1%; Pred. No. 4.3e-41;
Matches 352; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

```
QY 733 ACTCATTTCTCTGAGAAACCCACAGTGTATTTGTGAGAAAATGTTTCAACCGGAA 792
Db 475 ACCCATTTCTCCAGAAATCTCACCAGTGTGTCTACTGTGAGAGACGTTCAACCGGAA 416
QY 793 GATCATCTGAAGAATCAGTCTCATACACAGCCTTACAAAGAGACGTTTAAAGTGGAA 852
Db 415 GACCACTTGAAGAACCTCCAGACCCAGACCCCAACAAATGGCCTTTGGGTGTGAG 356
QY 853 GAATGTGGCAAGAACTACAATACCAAGCTTGGATTTAAACGTCCTTGGCCTTGCATGCC 912
Db 355 GAGTGTGGGAAGAGACTACAACCATGTCTGGCTATAAGAGGCACCTGGCCCTCCATGCG 296
QY 913 GCAACAGTGTGACCTCAGCTGTAAAGTATGTTTGCMAACTTTTGAAGCACGGGAGTG 972
Db 295 GCCAGCAGTGGGACCTCACCTGTGGGTCTGTGCGCTAGGAGCACCAGGAGTG 236
QY 973 CTTCCTGGAGCACCCTTAATCTCATGC---AGGCAAGTGTGTGGTGGGTTAAAGAAAA 1029
Db 235 CTACTGGACACCTCAAGCCCATGCGGAAGAGAGGCCCTTAGCGGAACCAAGGAAAAAG 176
QY 1030 AAGCACAGTGCAGAACTTGTGATGCGCGGTTCTACACCCGAAAGGATGCCGGAGACAC 1089
Db 175 AAGCACAGTGCAGACCACTGTGAAGATGCTTCTACACCCGGAAGGATGCGGACGCCAC 116
QY 1090 ATGGTGTGACACCTGGAGAAAGACTTCTCTGTGATTTGTGCACAGAGATTGGG 1149
Db 115 CTGGTGTCCACACAGGATGCAAGGACTTCTGTGCCAGTTCTGTGCCAGAGATTGGG 56
QY 1150 CGAAGGATCAGCTGACTCGACATATGAGNAGAGTCAATCAAGAGCTTCTGA 1204
Db 55 CGCAAGGATCACCTCACCGGCATACCAAGAAAGACCCACTCACAGGAGCTGATGA 1
```

Search completed: January 31, 2004, 06:29:02
Job time : 1686 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 06:29:05 ; Search time 13516 Seconds
(without alignments)
13150.239 Million cell updates/sec

Title: US-09-242-772-116
Perfect score: 7313
Sequence: 1 ggcagcgcatacactacaat.....tatgaataaaatctcgtgcc 7313

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	806.4	11.0	1201	13	BX424854	BX424854
2	793.6	10.9	866	13	BX370812	BX370812
3	758	10.4	909	13	BUI181917	BUI181917
4	704.4	9.6	734	14	CA411972	CA411972 UI-H-E20-

5	688	9.4	790	9	AUI138448
6	687	9.4	742	14	CA413047
7	685	9.4	693	12	BM724590
8	666	9.1	1201	13	BX402039
9	662.2	9.1	736	9	AL0366879
10	629	8.6	907	10	BG613583
11	590.8	8.1	616	13	BUE68683
12	570	7.8	570	9	AW339499
13	565.4	7.7	922	10	BG253502
14	556.4	7.6	659	29	AG152578
15	549.8	7.5	673	29	AG120959
16	540	7.4	540	9	AA706806
17	533.6	7.3	533.6	7.3	533.6
18	529.4	7.2	560	9	AI808840
19	518	7.1	518	10	BF115791
20	514	7.0	818	10	BF243649
21	505.4	6.9	597	10	AW977860
22	504	6.9	723	29	BZ845384
23	498.4	6.8	500	10	BF507694
24	494.2	6.8	719	29	BZ922587
25	482	6.6	482	9	AA707386
26	482	6.6	494	9	AW007402
27	464	6.3	466	9	AA628501
28	462.2	6.3	467	9	AW274762
29	448.2	6.1	458	9	AW002999
30	439	6.0	449	14	CA771364
31	438.2	6.0	454	9	AA418251
32	437.6	6.0	608	10	BG378156
33	434.2	5.9	683	14	BY719727
34	434	5.9	492	28	B94684
35	430	5.9	432	10	BF432133
36	427.4	5.8	460	9	AV745206
37	423	5.8	423	9	AI334157
38	422.2	5.8	427	9	AI223324
39	421.6	5.8	642	13	B0210438
40	417.2	5.7	421	14	D59273
41	406.6	5.6	489	9	AV745506
42	400.8	5.5	481	10	BF710381
43	391.2	5.3	4767	11	AK040626
44	388	5.3	522	12	BI298426
45	373	5.1	373	9	AA418372

ALIGNMENTS

RESULT 1
BX424854

LOCUS
DEFINITION

BX424854 Homo sapiens

XCL0BB001ZD02 5-PRIME, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1201)

Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 4473.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=XCL0BB001ZD02RP1&cluster=4473.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

1201 bp mRNA linear EST 13-MAY-2003
Homo sapiens cdna clone

AUI138448
CA413047 UI-H-E20-
BM724590 UI-E-E01-
BX402039 BX402039
AL036879 DKF2P564P
BG613583 603641935
BUE68683 UI-CF-DU1
AW339499 xz91905.x
BG253502 602364227
AG152578 Pan trogl
AG120959 Pan trogl
AA706806 zj30c09.s
BY734789 BY734789
A1808840 wf97h04.x
BF115791 7h64h11.x
BF243649 601878620
AW977860 EST389864
BZ845384 CH240_250
BF507694 UI-H-BW1-
BZ922587 CH240_121
AA707386 zj27e09.s
AM007402 wt55b04.x
AA628501 af27d06.s
AW274762 xn34h01.x
AW002999 wq2d08.x
CA771364 1o70c07.y
AA418251 zv96c12.s
BG378156 UI-R-CV1-
BY719727 BY719727
B94684 CIT-HSP-217
BR432133 nab79501.
AV745206 AV745206
AI334157 QG07h12.x
AI223324 QG71c02.x
BU210438 604156514
D59273 HUM007A10A
AV745506 AV745506
BF710381 MI-P-AV1-
AK040626 Mus muscu
BI298426 UI-R-CV2-
AA418372 zv96g12.r

1493 QY CTACCTCATATGCAATTTCTATTTCTGAAAGAACACAGCCATTAAGGGGAAATGAGA 1552
1088 Db CTACCTCATATGCAATTTCTATTTCTGAAAGAACACAGCCATTAAGGGGAAATGAGA 1145
1553 QY GTTACCTGATGAGTTACAAGTGGCGTCCCTCTTCTATCCCAAGATTCTCAAGATC 1610
1146 Db RKTACCGGATGAGTTACAAGKGGC-TKCCCTCTCATCCC-ARATWCTCAAGCATS 1201
RESULT 2
BX370812 866 bp mRNA linear EST 08-MAY-2003
BX370812 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL012YM24 5-PRIME, mRNA sequence.
ACCESSION BX370812
VERSION BX370812.1 GI:30445558
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 866)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4473.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG062D02_CS05868_1&cluster=4473.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0BAG062D02_CS05868_1.
FEATURES
Location/Qualifiers
1..866
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL012YM24"
/cell_type="B CELLS (RAMOS CELL LINE)"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 235 a 218 c 177 g 236 t
ORIGIN
Query Match 10.9%; Score 793.6; DB 13; Length 866;
Best Local Similarity 98.5%; Pred. No. 3.1e-72; Indels 4; Gaps 4;
Matches 843; Conservative 0; Mismatches 9;
QY 984 CCTTAATCTCATGAGGCAAGTCTCTGTGGGTAAAGAAACCAACGACGATGCGA 1043
Db 14 CCTTAATCTCATGAGGCAAGTCTCTGTGGGTAAAGAAACCAACGACGATGCGA 71
QY 1044 ACATTGTGATCGCGGTTCTACACCGAAGATGTCCGAGACATGTTGGCGGAAAGGATCACCT 1103
Db 72 ACATTGTGATCGCGGTTCTACACCGAAGATGTCCGAGACATGTTGGCGGAAAGGATCACCT 130
QY 1104 TGAAGAAAGGACTTCTCTGTGATGTGACAGATTTGGCGGAAAGGATCACCT 1163
Db 131 TGAAGAAAGGACTTCTCTGTGATGTGACAGATTTGGCGGAAAGGATCACCT 190
QY 1164 GACTCCACATATCAAGAAAGTCAATCAAGAGCTTCTGAAGCTCAAAACAGAACAGT 1223
Db 191 GACTCCACATATCAAGAAAGTCAATCAAGAGCTTCTGAAGCTCAAAACAGAACAGT 250

Paradise Avenue Genoscope sequence ID : XCL0BB0012D02RP1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="XCL0BB0012D02"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 354 a 284 c 274 g 271 t 18 others
ORIGIN
Query Match 11.0%; Score 806.4; DB 13; Length 1201;
Best Local Similarity 97.0%; Pred. No. 1.3e-73; Indels 7; Gaps 6;
Matches 871; Conservative 12; Mismatches 8;
QY 714 ATTCAAGGCACATGCTACTCTCTCTGAGAAACCCCAAGTGTAAATTATTGTGA 773
Db 310 AGTAATCAGGCATGCTACTCTCTCTGAGAAACCCCAAGTGTAAATTATTGTGA 369
QY 774 GAAATGTTTACCGGAAAGATCATCTGAAAGATCACTCATACACGACCTTAACAA 833
Db 370 GAAATGTTTACCGGAAAGATCATCTGAAAGATCACTCATACACGACCTTAACAA 429
QY 834 AGAGACGTTTAAAGTGGCAAGATGTGCAAGAACTACAATCAAGCTTGGATTAAACG 893
Db 430 AGAGACGTTTAAAGTGGCAAGATGTGCAAGAACTACAATCAAGCTTGGATTAAACG 489
QY 894 TCATTGGCCTTGATCGGCAAGTGGTGAAGTCACTCACTGTAAAGTATTTGCAAAC 953
Db 490 TCATTGGCCTTGATCGGCAAGTGGTGAAGTCACTCACTGTAA- GTAATGTTGCAAAC 548
QY 954 TTTTGAAGACGCGGAGTCTCTGGAGCCTTTAAATCTCATGAGGCAAGTCTGTGG 1013
Db 549 TTTTGAAGACGCGGAGT-CTTCTGGAGCCTTTAAATCTCATGAGGCAAGTCTGTGG 607
QY 1014 TGGGGTTAAAGAAAGAACAGCAGTGGCAATGTGATCGCGGTTCTACACCGAAA 1073
Db 608 TGGGGTTAAAGAAAGAACAGCAGTGGCAATGTGATCGCGGTTCTACACCGAAA 667
QY 1074 GGATGTCGGAGACACATGTTGGTGCACACTGGAAGAAAGGACTTCTCTGTGATTTG 1133
Db 668 GGATGTCGGAGACACATGTTGGTGCACACTGGAAGAAAGGACTTCTCTGTGATTTG 727
QY 1134 TGCACAGATTTGGCGGAAAGGATCACTGACTCGAATATGAAGAGAGTCAATCA 1193
Db 728 TGCACAGATTTGGCGGAAAGGATCACTGACTCGAATATGAAGAGAGTCAATCA 787
QY 1194 AGAGCTTCTGAAGTCAAAACAGACAGTGAATTTCTTGACCCATTTACCTGCAATGT 1253
Db 788 AGAGCTTCTGAAGTCAAAACAGACAGTGAATTTCTTGACCCATTTACCTGCAATGT 847
QY 1254 GTCTGTGCTATAAAGACGAGCTCTCTCGGTGATGTCTTACCTTCCAGTGAATGTT 1313
Db 848 GTCTGTGCTATAAAGACGAGCTCTCTCGGTGATGTCTTACCTTCCAGTGAATGTT 907
QY 1314 ATCAAGGCCATTACAAACACTTTGAGTTAAACCTCTCAACACTCCGATTCAGTCCAT 1373
Db 908 ATCAAGGCCATTACAAACACTTTGAGTTAAACCTCTCAACACTCCGATTCAGTCCAT 967
QY 1374 GCAGAGCTGGGATCTGCCCAACAAATGATCAACTTTACCTTTGGGATGACATGCC 1433
Db 968 GCAGAGCTGGGATCTGCCCAACAAATGATCAACTTTACCTTTGGGATGACATGCC 1027
QY 1434 AATAGATGAGACACTGTTTCATCCC-TCTCACCACCTTTCTTTCAAATATCCGTTCA 1492
Db 1028 AATAGATGAGACACTGTTTCATCCC-TCTCACCACCTTTCTTTCAAATATCCGTTCA 1087


```

QY 1224 GGAATTCCTGACCCATTTACTGCAATGTGTCTGTGCTATATAAGAGAGAGCTCTTCC 1283
Db |||||||
QY 1251 GGAATTCCTGACCCATTTACTGCAATGTGTCTGTGCTATATAAGAGAGAGCTCTTCC 310
Db |||||||
QY 1284 GGTGATGTCCTTACCTTCAGTGAACCTGTATCAAGGCATTCACAAACATTTGCGATT 1343
Db |||||||
QY 1311 GGTGATGTCCTTACCTTCAGTGAACCTGTATCAAGGCATTCACAAACATTTGCGATT 370
Db |||||||
QY 1344 AAACCTCTACAACTCCATTTCACTGATCCAGTCGAGAGCTCGGATCTGCCACCAAAATGAT 1403
Db |||||||
QY 1371 AAACCTCTACAACTCCATTTCACTGATCCAGTCGAGAGCTCGGATCTGCCACCAAAATGAT 430
Db |||||||
QY 1404 CACAACCTTACCTTTGGGAATGACATGCCCAATAGATATGGACACTGTTTCATCCCTCTCA 1463
Db |||||||
QY 1431 CACAACCTTACCTTTGGGAATGACATGCCCAATAGATATGGACACTGTTTCATCCCTCTCA 490
Db |||||||
QY 1464 CCACCTTCTCTTCAAAATATCGGTTTCAGTTCCTACCTCATATGCAATTTCTATTCCTGAAA 1523
Db |||||||
QY 1491 CCACCTTCTCTTCAAAATATCGGTTTCAGTTCCTACCTCATATGCAATTTCTATTCCTGAAA 550
Db |||||||
QY 1524 AGACACGCAATTAAGGGGAAATTTGAGAGTTACCTGATGGAGTTACAAGTGCGCGTCC 1583
Db |||||||
QY 1551 AGACACGCAATTAAGGGGAAATTTGAGAGTTACCTGATGGAGTTACAAGTGCGCGTCC 610
Db |||||||
QY 1584 CTCTTCATCCCAAGATCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCCTCAGAT 1643
Db |||||||
QY 1611 CTCTTCATCCCAAGATCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCCTCAGAT 670
Db |||||||
QY 1644 TGGGTCCTTAGATGATGGTGAGGAGACCTCTCCCTATCCAAAGCTCTATCTCCATCAG 1703
Db |||||||
QY 1671 TGGGTCCTTAGATGATGGTGAGGAGACCTCTCCCTATCCAAAGCTCTATCTCCATCAG 730
Db |||||||
QY 1704 TGACCCCTTAAACACACAGCATTTGATTTTCTCAGTTGTTTAAATTTTCATACCTTTAA 1763
Db |||||||
QY 1731 TGACCCCTTAAACACACAGCATTTGATTTTCTCAGTTGTTTAAATTTTCATACCTTTAA 790
Db |||||||
QY 1764 TGGGTCCTTAGATGATGGTGAGGAGACCTCTCCCTATCCAAAGCTCTATCTCCATCAG 1823
Db |||||||
QY 1791 TGGGTCCTTAGATGATGGTGAGGAGACCTCTCCCTATCCAAAGCTCTATCTCCATCAG 849
Db |||||||
QY 1824 AGCACATTCCTCTGTT 1839
Db |||||||
QY 850 AGCACATTCCTCTGTT 865
Db |||||||

RESULT 3
BUI81917
LOCUS
DEFINITION AGENCOURT_7780674 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6137707
5', mRNA sequence.
ACCESSION BUI81917
VERSION BUI81917.1 GI:22695901
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 909)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: csaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13451 row: o column: 20
High quality sequence stop: 683.

```

FEATURES

```

source
Location/Qualifiers
1..909
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:6137707"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 329 a 132 c 142 g 305 t 1 others
ORIGIN
Query Match 10.4%; Score 758; DB 13; Length 909;
Best Local Similarity 95.8%; Pred. No. 1.3e-68;
Matches 844; Conservative 0; Mismatches 26; Indels 11; Gaps 6;
QY 3255 ATTAGAAAACCTACCTGATGAACCTACCAAAAGTTTAAAGTGACTAGAAATATATACAGTA 3314
Db |||||||
QY 3315 AAATCCCACTTTCATGCACTCTCGGAAATGATAGGAGTATTCGAAATAGTTGAGTTTG 3374
Db |||||||
QY 3375 TAGAGGGTAAACAAAGTAAAGTAAACAAACCTATCTTGGTTAAACATGAAATTAACATTTG 3434
Db |||||||
QY 3435 AGAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 180
Db |||||||
QY 3495 ATCTTCTTAAAGGTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3554
Db |||||||
QY 3555 AAACATATCTATTAAGCACTTGTAAACACCTTATTTAGGACCTTTTCGTTGGGATGG 3614
Db |||||||
QY 3615 GGGCAAGGTGGGAGTTTGTAGAGAGTATATATCTCTTTTAAACACCTTATTTAGGACCTTT 3673
Db |||||||
QY 3673 GGGCAAGGTGGGAGTTTGTAGAGAGTATATATCTCTTTTAAACACCTTATTTAGGACCTTT 420
Db |||||||
QY 3674 AAATATTTCTGAGCACTCAATTAGCCCTATATGAAACTTCTTCTTTTAAACACCTTATTT 3733
Db |||||||
QY 3733 AAATATTTCTGAGCACTCAATTAGCCCTATATGAAACTTCTTCTTTTAAACACCTTATTT 480
Db |||||||
QY 3734 TTATCACTGAGATTCGAATGTTTACCAAGATTTCTTAAACATGAGTCAGATTACTGAA 3793
Db |||||||
QY 3794 TATAATACATTTTAAATATTTTGGGAGTAGTATATTTTGTGAGAAATGTAATTTGTA 540
Db |||||||
QY 541 TATAATACATTTTAAATATTTTGGGAGTAGTATATTTTGTGAGAAATGTAATTTGTA 3853
Db |||||||
QY 3854 ATATGTAAATGGGGCTTCAATATATATATATATATATATATATATATATATATATATAT 600
Db |||||||
QY 601 ATATGTAAATGGGGCTTCAATATATATATATATATATATATATATATATATATATATAT 3913
Db |||||||
QY 3914 CATACCGCACTTCATAGATCAAAAGTTGCTCTGGAAGGAGCTTTGGCTCTCTGATATTTT 659
Db |||||||
QY 660 CATACCGCACTTCATAGATCAAAAGTTGCTCTGGAAGGAGCTTTGGCTCTCTGATATTTT 3973
Db |||||||
QY 3974 ATCATGCTCTATATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAATAT 719
Db |||||||
QY 4032 ATCATGCTCTATATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAATAT 4032
Db |||||||
QY 4033 A--TTTATATGAAAAATTAACATTTTATTAATAA--GCTGCTTCCAAAGCAATTAATA 779
Db |||||||

```

734 TGACATGCCAATAGATATGGACACTGTTTCATCCCTCTCACACCTTCTTTCAATATC 675
1484 CGTTCAGTTTACCTCATATGCAATTTCTATTCTCTGAAAAAGAACAGCATTAAGGG-G 1542
674 CGTTCAGTTTACCTCATATGCAATTTCTATTCTCTGAAAAAGAACAGCATTAAGGGNN 615
1543 GAAATTGAGAGTTTACCTCATATGCAATTTCTATTCTCTGAAAAAGAACAGCATTAAGGGNN 1602
614 GAAATTGAGAGTTTACCTCATATGCAATTTCTATTCTCTGAAAAAGAACAGCATTAAGGGNN 555
1603 CAAGCATCGTATCATATGCAATTTCTATTCTCTGAAAAAGAACAGCATTAAGGGNN 1662
554 CAAGCATCGTATCATATGCAATTTCTATTCTCTGAAAAAGAACAGCATTAAGGGNN 495
1663 GCAGGAGACCTCTCCCTATCCAAAGCTCTATCTCCATCAGTGCACCCCTTAAACACACCA 1722
494 GCAGGAGACCTCTCCCTATCCAAAGCTCTATCTCCATCAGTGCACCCCTTAAACACACCA 435
1723 GCATTGGATTTTCTCAGTTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 1782
434 GCATTGGATTTTCTCAGTTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 375
1783 CTATCAGTGGGAGCCCTTGGAAATGAGCTATTCCTCAGTGCACCCCTTAAACACACCA 1842
374 CTATCAGTGGGAGCCCTTGGAAATGAGCTATTCCTCAGTGCACCCCTTAAACACACCA 315
1843 CAGTCTCCCAACACACACACAGGATCTTCAGGATCTTCAGGATCTTCAGGATCTTCAGGATCTTC 1902
314 CAGTCTCCCAACACACACACAGGATCTTCAGGATCTTCAGGATCTTCAGGATCTTCAGGATCTTC 255
1903 CTGCACTCACTGTGAGCAGCTTTTACAGCAGTTCCTCAGGATCTTCAGGATCTTCAGGATCTTC 1962
254 CTGCACTCACTGTGAGCAGCTTTTACAGCAGTTCCTCAGGATCTTCAGGATCTTCAGGATCTTC 195
1963 TTCCATCAAGCTTTTCTCAGTGGATCTTCAGGATCTTCAGGATCTTCAGGATCTTCAGGATCTTC 2022
194 TTCCATCAAGCTTTTCTCAGTGGATCTTCAGGATCTTCAGGATCTTCAGGATCTTCAGGATCTTC 135
2023 GCTGTGCCCTAGATGACCATTTTATTTTATTTAGTGGCTACTTTTAAACACAGTATATAAATTC 2082
134 GCTGTGCCCTAGATGACCATTTTATTTTATTTAGTGGCTACTTTTAAACACAGTATATAAATTC 75
2083 TGCTTTTGTATAATACAAATTTTCAATTAAGCAGTATATAAATAGAACTAGCTTTTAA 2141
74 TGCCTTTGTATAATACAAATTTTCAATTAAGCAGTATATAAATAGAACTAGCTTTTAA 16

780 AATTATATGAAAAAATTAATTTTATTAATAAAGTGTGTTTCAAAAGCCATTAATA 839
4088 ---TTATATATGTTAATAGGAGTACATTTTAAATTTT 4125
840 ATTATATATGTTAATAAAGGAAAGTACCTTTTAAATTTT 880

CA411972 734 bp mRNA linear EST 07-NOV-2002
UI-H-EZO-bal-m-01-0-UI.s1 NCI CGAP Chl Homo sapiens cDNA clone
UI-H-EZO-bal-m-01-0-UI 3', mRNA sequence.

CA411972 1 GI:24774623
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 734)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: gsapbs-r@mail.nih.gov
Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of
Orthopedics
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-22, >AT rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source Location/Qualifiers
1..734
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-EZO-bal-m-01-0-UI"
/tissue_type="Chondrosarcoma Grade II"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Chl"
/note="Organ: Left Pelvis; Vector: pT73-pac (Pharmacia)
with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP Chl is a cDNA library containing the following
tissue(s): Chondrosarcoma Grade II. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
TGATCACGCT.
TAG LIB=UI-H-EZO
TAG TISSUE=grade-2-chondrosarcoma
TAG SEQ=ATCAATATG"

BASE COUNT 216 a 123 c 179 g 214 t 2 others
ORIGIN

Query Match 9.6%; Score 704.4; DB 14; Length 734;
Best Local Similarity 99.6%; Pred. No. 4.4e-63;
Matches 716; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1424 TGACATGCCAATAGATATGGACACTGTTTCATCCCTCTCACACCTTCTTTCAATATC 1483

HRI human cDNA project; 5'- and 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES

source Location/Qualifiers

RESULT 6

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BASE COUNT 241 a 138 c 140 g 267 t 4 others

Query Match 9.4%; Score 688; DB 9; Length 790; Best Local Similarity 95.8%; Pred. No. 2e-61; Mismatches 0; Gaps 5; Matches 759; Conservative 0; Indels 24; Indels 9; Gaps 5;

QY 2190 GTCCGATTCCTGCTGCAATTCACATTTTAAAGACTTTTTCATATAGGAAGC 2249
Db 1 GTCCGATTCCTGCTGCAATTCACATTTTAAAGACTTTTCATATAGGAAGC 57QY 2250 CATTATTAGTAGTAACCTTTTACAAATCCCATTTTCAAAATTTAGATCTTAAATTT 2309
Db 58 CATTATTAGTAGTAACCTTTTACAAATCCCATTTTCAAAATTTAGATCTTAAATTT 117QY 2310 TTCAATTTTGTCTAATAACAGTGGCTCTACCTTTTGACATCTGCTCATTTAAATTTA 2369
Db 118 TTCAATTTTGTCTAATAACAGTGGCTCTACCTTTTGACATCTGCTCATTTAAATTTA 177QY 2370 GCAATAGAAATGTAATTTGTAATAAAGTTTGTGAATAACTCAAGGGTTTAAATTTCTTA 2429
Db 178 GCAATAGAAATGTAATTTGTAATAAAGTTTGTGAATAACTCAAGGGTTTAAATTTCTTA 237QY 2430 CTAGCTTTCAAAATGGATTAATTAATCAAGTCTTCAATGAATTAAGATCCAGTTTCGGA 2489
Db 238 CTAGCTTTCAAAATGGATTAATTAATCAAGTCTTCAATGAATTAAGATCCAGTTTCGGA 297QY 2490 AGATAATAAATGTTGTTAGATACACCAATTAATTCAGATCAGTATATTCGGAAGCTCTC 2549
Db 298 AGATAATAAATGTTGTTAGATACACCAATTAATTCAGATCAGTATATTCGGAAGCTCTC 357QY 2550 TGTGTTCTGGCTAAATATTTGTCATCTTTATATGAGCTTTTAAAGGAAACAAACCCCTA 2609
Db 358 TGTGTTCTGGCTAAATATTTGTCATCTTTATATGAGCTTTTAAAGGAAACAAACCCCTA 417QY 2610 AACACAAAGCATCAGTATTTATAGCAAAAGAGACTCTGTTAGGTGACATGGCATTTGCT 2669
Db 418 AACACAAAGCATCAGTATTTATAGCAAAAGAGACTCTGTTAGGTGACATGGCATTTGCT 477QY 2670 GTCACTTAATAGTTGGCCCTAAATTTAGTACACAGATATTTGTCGTTTCATCTTCT 2729
Db 478 GTCACTTAATAGTTGGCCCTAAATTTAGTACACAGATATTTGTCGTTTCATCTTCT 537QY 2730 TAAATGCTATCTTTTCAATTAATAT -AGTAATAGTGTATGGCATTTGAGTCTTCAGAG 2788
Db 538 TAAATGCTATCTTTTCAATTAATAT -AGTAATAGTGTATGGCATTTGAGTCTTCAGAG 597QY 2789 TCGATATATAGTAGATCTCTTTAGTCTTTTCCACCTTTCAATCCAAAGGGGTGGTCAA 2848
Db 598 TCGATATATAGTAGATCTCTTTAGTCTTTTCCACCTTTCAATCCAAAGGGGTGGTCAA 657QY 2849 GTGCAGCCAGCAATTTATTTTTCATTTGTCGCCACGGTTAGTCCATATCTAGAGCCATT 2908
Db 658 GTGCAGCCAGCAATTTATTTTTCATTTGTCGCCACGGTTAGTCCATATCTAGAGCCATT 716QY 2909 GTGGAATGCGAGCATGAGGTGTTTATCCACAGTGGATTGACTCAGCTCTGTGGGT 2968
Db 717 GTGGAATGCGAGCATGAGGTGTTTATCCACAGTGGATTGACTCAGCTCTGTGGGT 772

QY 2969 GACAGACTTCTA 2980

Db 773 ACAAACTTCAA 784

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CA413047

UI-H-EZ0-bap-e-03-0-UI.s1 NCI CGAP Chl Homo sapiens cDNA clone

CA413047

CA413047.1 GI:24775698

EST.

Homo sapiens (human)

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 742)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaaps-remail.nih.gov

Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of

Orthopaedics

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained

from Dr. M. Bento Soares, bento-soares@uiowa.edu

The following repetitive elements were found in this cDNA

sequence: 1-21, >AT richLow_complexity (matched complement)

Seq primer: M13 FORWARD

POLYA=Yes.

Location/Qualifiers

1. .742

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-H-EZ0-bap-e-03-0-UI"

/tissue_type="Chondrosarcoma Grade II"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies)"

/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)

with a modified polylinker; Site 1: EcoR I; Site 2: Not I;

NCI-CGAP Chl is a cDNA library containing the following

tissue(s): Chondrosarcoma Grade II. The library was

constructed according to Bonaldo, Lennon and Soares,

Genome Research, 6:791-806, 1996. First strand cDNA

synthesis was primed with an oligo-dr primer containing a

Not I site. Double stranded cDNA was ligated to an EcoR I

adaptor, digested with Not I, and cloned directionally

into pT73-Pac vector. The oligonucleotide used to prime

the synthesis of first-strand cDNA contains a library tag

sequence that is located between the Not I site and the

(dT)18 tail. The sequence tag for this library is

TGATCAGCT.

TAG LIB=UI-H-EZ0

TAG TISSUE=grade-2-chondrosarcoma

TAG SEQ=ATCTAATATG"

BASE COUNT 236 a 129 c 102 g 274 t 1 others

ORIGIN

Query Match 9.4%; Score 687; DB 14; Length 742;

Best Local Similarity 99.7%; Pred. No. 2.6e-61;

Matches 709; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 6596 GTATTGATGCTTGCACAAAGAGATCTCGGACATGCAAAATGTTACCAACAAACCC 6655

Db 742 GTATTGATGCTTGCACAAAGAGATCTCGGACATGCAAAATG-TACCAGACAAACCC 684

QY 6656 AGCTTTGTCCAC-AAGGTGACTGTAACTCAGAAAGTGGCTTTTATAATAGGGTG 6714

DNA sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES
SOURCE

683	Db	AGCTTTTGTCCACNAGGTGACTGTAACTCAGAAATGGAAGTGGGCTTTTAAATAGGGTG	624
6715	QY	TGGAGTGAAGAACATGCTGTATGTACTAAACAGCCCTTTGAAATTTAAACAAAACTGGAA	6774
623	Db	TGGAGTGAAGAACATGCTGTATGTACTAAACAGCCCTTTGAAATTTAAACAAAACTGGAA	564
6775	QY	TCCATTAGGAAACGGATTCATCATACCTGAAACATAAGCTGACCTGCGAAATTTGATTT	6834
563	Db	TCCATTAGGAAACGGATTCATCATACCTGAAACATAAGCTGACCTGCGAAATTTGATTT	504
6835	QY	TTAGCTAATGAAAAAGTGTTTCGAGCTAGTACCTAAATAATGTTCTAAATGATAAAGTTTG	6894
503	Db	TTAGCTAATGAAAAAGTGTTTCGAGCTAGTACCTAAATAATGTTCTAAATGATAAAGTTTG	444
6895	QY	AGTCAAATAGAAAAAGAAAAAANTCTGCATTCAGGCCGAATTTTGTATATTTTATTCG	6954
443	Db	AGTCAAATAGAAAAAGAAAAAANTCTGCATTCAGGCCGAATTTTGTATATTTTATTCG	384
6955	QY	ATTTAAAAATTGCTATTCTGTAAATTTGGGAAATCAAAGTGGCTTTATCATGTATATCGTGA	7014
383	Db	ATTTAAAAATTGCTATTCTGTAAATTTGGGAAATCAAAGTGGCTTTATCATGTATATCGTGA	324
7015	QY	CTTAAAAATGATTTCACAAACTACTGTTGTATTTGTTATAAATATAGACAAAGATCATATT	7074
323	Db	CTTAAAAATGATTTCACAAACTACTGTTGTATTTGTTATAAATATAGACAAAGATCATATT	264
7075	QY	TTTGTGTGTGTATAAGCTCTGTAAAAATAGCAATCACAATTATGAAGCTGCAGTGACTA	7134
263	Db	TTTTTGTGTGTATAGCTCTGTAAAAATAGCAATCACAATTATGAAGCTGCAGTGACTA	204
7135	QY	CATTTTAAACATTCACATCCAAAGACAGACTATTTTGTCCATATACCAAGATTTAAA	7194
203	Db	CATTTTAAACATTCACATCCAAAGACAGACTATTTTGTCCATATACCAAGATTTAAA	144
7195	QY	ATATTAAATTTGCTCTAATTAATAAATAGTACTCGAGCTTCTTGTGCCCTACAGTGTTAT	7254
143	Db	ATATTAAATTTGCTCTAATTAATAAATAGTACTCGAGCTTCTTGTGCCCTACAGTGTTAT	84
7255	QY	GTTTGCCTGAAGATATAGTATGTAATTCACAAAAATATATGATATAAAT	7305
83	Db	GTTTGCCTGAAGATATAGATATGTAATTCACAAAAATATATGATATAAAT	33

RESULT 7	
BM724590	693 bp mRNA linear EST 01-MAR-2002
LOCUS	
DEFINITION	UI-E-E01-aiz-j-24-0-UI_r1 UI-E-E01 Homo sapiens cDNA clone
ACCESSION	UI-E-E01-aiz-j-24-0-UI 5', mRNA sequence.
VERSION	
KEYWORDS	
SOURCE	BM724590.1 GI:19045921
ORGANISM	EST.
	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 693)
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
PUBMED	8889548
COMMENT	Contact: Soares, MB Coordinated Laboratory for Computational Genomics University of Iowa 375 Newton Road, 4156 MEERF, Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: bento-soares@uiowa.edu Tissue Procurement: Dr. Gregg Hageman cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrived by: Dr. M. Bento Soares, University of Iowa

DEFINITION DKFZp564P1863_r1 564 (synonym: hfb2) Homo sapiens cDNA clone
ACCESSION DKFZp564P1863 5', mRNA sequence.
VERSION AL036879 GI:5927934
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 736)
JOURNAL Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann
COMMENT EST (Duesterhoeft, et al.)
Unpublished
On Jul 7, 1999 this sequence version replaced gi:5866332.
Contact: Duesterhoeft A
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Olgan (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp564P1863) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Location/Qualifiers
1..736
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp564P1863"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="xl-2blue"
/clone_lib="564 (synonym: hfb2)"
/note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"
BASE COUNT 223 a 174 c 157 g 182 t
ORIGIN
Query Match 9.1%; Score 662.2; DB 9; Length 736;
Best Local Similarity 98.1%; Pred. No. 9e-59;
Matches 723; Conservative 0; Mismatches 8; Indels 6; Gaps 5;
759 GTGTAATTATGCGAAGAAATGTTTCACCGGAAAGATCATCTGAAGATCACTCCATAC 818
2 GGGTAATTATGCGAAGAAATGTTTCACCGGAAAGATCATCTGAAGATCACTCCATAC 61
819 ACACGACCCCTAACAAAGACAGCGTTTAAAGTGGCAAGATGCGCAAGCACTCAATACCAA 878
62 ACACGACCCCTAACAAAGACAGCGTTTAAAGTGGCAAGATGCGCAAGCACTCAATACCAA 121
879 GCTTGATTTAAACGTCACCTGCGCTTGCATGCGCAACAAAGTGGTCACTCACTGTAA 938
122 GCTTGATTTAAACGTCACCTGCGCTTGCATGCGCAACAAAGTGGTCACTCACTGTAA 181
939 GGTATGTTGCAAACTTTTGAAGACACGGAGTCTTCTGGAGCAGCTTAAATCTCATGC 998
182 GGTATGTTGCAAACTTTTGAAGACACGGAGTCTTCTGGAGCAGCTTAAATCTCATGC 241
999 AGCGAAGTCGTCGTGGGGTTTAAAGAAAAAGACCACTGCGCAAGATGTCATGCGCG 1058
242 AGCGAAGTCGTCGTGGGGTTTAAAGAAAAAGACCACTGCGCAAGATGTCATGCGCG 301
1059 GTTCTACACCGGAAGATGTCGCGAGACACATGGTGGTGCACACTGGAAGAAAGACTT 1118
302 GTTCTACACCGGAAGATGTCGCGAGACACATGGTGGTGCACACTGGAAGAAAGACTT 361
1119 CCTCTGTCAGTATTGTGCACAGAGATTGGCGGAAAGGATCACTGATGCATATGAA 1178
362 CCTCTGTCAGTATTGTGCACAGAGATTGGCGGAAAGGATCACTGATGCATATGAA 421

1179 GAAGAGTCACATCAAGAGCTTCTGAAGCTCAAAACAGAACCAAGTGGATTCTCTTGACCC 1238
422 GAAGAGTCACATCAAGAGCTTCTGAAGCTCAAAACAGAACCAAGTGGATTCTCTTGACCC 481
1239 ATTTACCTGCAATGTGTCTGTGCTCTATATAAAGACGAGCTCTCTCCGGTGTATGCTTACC 1298
482 ATTTACCTGCAATGTGTCTGTGCTCTATATAAAGACGAGCTCTCTCCGGTGTATGCTTACC 541
1299 TTCCAGTG- AACTGTTTATCAAGGCCATTCAAAACACTTTGCAAGTTAAACCTCTACAACA 1357
542 TTCCAGTGAACTGTTTATCAAGGCCATTCAAAACACTTTGCAAGTTAAACCTCTACAACA 601
1358 CTCATTTCAGTCCATGCG-AGAGCTCGGATCTGCCAC- AAATGATCAACAACCTTTACC 1415
602 CTCATTTCAGTCCATGCGAGAGCTCGGATCTGCCAC- AAATGATCAACAACCTTTACC 661
1416 TTTGGGAATGACATGCCCAATAGATAT- GGACACTGTTCACTCCCTCTCACACCTTTCTT 1474
662 TTTGGGAATGACATGCCCAATAGATAT- GGACACTGTTCACTCCCTCTCACACCTTTCTT 721
1475 TCAAATATCCGTTTCACT 1491
722 --CAATATCCGTTTCACT 736

RESULT 10
BG613583 907 bp mRNA linear EST 18-APR-2001
LOCUS 602641935P1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4772744 5',
DEFINITION mRNA sequence.
ACCESSION BG613583
VERSION BG613583.1 GI:13664954
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 907)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1643 row: f column: 09
High quality sequence stop: 645.
FEATURES
Location/Qualifiers
1..907
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4772744"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC 61"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccatattggc); Site_2: SfiI (ggccatattggc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor
sequence: 5'-ATTCGAGGCGGCGGCGGCGGAGATG-DT(30)BN-3'
(Where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 Kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library."

BASE COUNT	267 a	202 c	174 g	264 t
ORIGIN				
Query Match	8.6%; Score 629; DB 10; Length 907;			
Best Local Similarity	87.7%; Pred. No. 1.9e-55;			
Matches 774; Conservative	0; Mismatches 85; Indels 24; Gaps 7;			
QY	2763 AGTGATGTCATTTGGGCTTCAGAGTCGATATATAGGTAGATCTCTTTAGTCTTTTCCA 2822			
Db	1 AGTGATGTCATTTGGGCTTCAGAGTCGATATATAGGTAGATCTCTTTAGTCTTTTCCA 60			
QY	2823 CTTTTCACATCCAAAGGGTGGGTCAAGTCGACCGACATTTATTTTCATTTGTCGCCA 2882			
Db	61 CTTTTCACATCCAAAGGGTGGGTCAAGTCGACCGACATTTATTTTCATTTGTCGCCA 120			
QY	2883 CGTTAGTCCATPAATCTAGAGCCATTTGGAACTGCGACATGAGTGTGTTTATCCAC 2942			
Db	121 CGTTAGTCCATPAATCTAGAGCCATTTGGAACTGCGACATGAGTGTGTTTATCCAC 180			
QY	2943 AGTGATGTCATGACCTCTGTCGGTGCAGACTTCTAAGCAGGAAGATAGACGTGAAGC 3002			
Db	181 AGTGATGTCATGACCTCTGTCGGTGCAGACTTCTAAGCAGGAAGATAGACGTGAAGC 240			
QY	3003 ACATGGTTACATTTGGGAATTTGTCAGGATCATGGCCCTGTAGCCAGGTTAAAAAC 3062			
Db	241 ACATGGTTACATTTGGGAATTTGTCAGGATCATGGCCCTGTAGCCAGGTTAAAAAC 299			
QY	3063 TGGACTTTTAAAGTAAAGTAAAGCATGCGCTTATATCATTTCTTGTGTAATTTGAT 3122			
Db	300 TGGACTTTTAAAGTAAAGTAAAGCATGCGCTTATATCATTTCTTGTGTAATTTGAT 359			
QY	3123 ATGTTTTTCTTCCCTTAAGATCAAGACAGAGAAAACAAAACACACGCTCTACCCGAT 3182			
Db	360 ATGTTTTTCTTCCCTTAAGATCAAGACAGAGAAAACAAAACACACGCTCTACCCGAT 419			
QY	3183 GTTATCTTCTGATTCATGATGATCCATCTTCTTCCTTGCATATTTTGGATGGAGATTT 3242			
Db	420 GTTATCTTCTGATTCATGATGATCCATCTTCTTCCTTGCATATTTTGGATGGAGATTT 479			
QY	3243 GAAGTTAAATGCAATTAAGAACTACCTGATGAACTACCAACAAAGTTTAAAGTGAAGTAA 3302			
Db	480 GAAGTTAAATGCAATTAAGAACTACCTGATGAACTACCAACAAAGTTTAAAGTGAAGTAA 539			
QY	3303 ATATATACAGTAAATCCCACTTTCAGCATCTCTGGAAATGATAGAGTATTTGCAAT 3362			
Db	540 ATATATACAGTAAATCCCACTTTCAGCATCTCTGGAAATGATAGAGTATTTGCAAT 599			
QY	3363 AAGTTGAGTTTGTAGAGGGTAAACAAAGTAAAGTAAACAACTATCTTGGTTAAAT-- 3420			
Db	600 AAGTTGAG--TTGTAGAGGTAAACAAAGTAAACAACTATCTTGGTTAAAT-- 3474			
QY	3421 -GAAATACAAATTGAGATATATT-ATATTCACTGAATAATTAT---AGGCTTTTCTT 3474			
Db	659 ACCAATTCACCAATTGAGATATCTTCATATCTACTGACTACTCTCTCGGCTTTTCT 718			
QY	3475 CACATTAGACACCAACATAATCTCTTAAAGT-----CTAATTAATATATTT 3523			
Db	719 CACATTAGACACCAACATAATCTCTTAAAGT-----CTAATTAATATATTT 3523			
QY	3524 TTCTAAGGTCAGTTGGGACATTAAC---CTAAGAAACATATCTATTAGCACTTTGTTAA 3580			
Db	779 TCCTAAGGTCAGTTGGGACATTAAC---CTAAGAAACATATCTATTAGCACTTTGTTAA 3580			
QY	3581 CACCTTATTTTAGGACCTTTCCGTTGGGATGGGGGCAAGG 3623			
Db	839 ACCTTAGTGTAGGACCTTTCCGTTGGGATGGGGGCAAGG 881			
RESULT 11				
BU686683/c				
LOCUS	616 bp mRNA linear EST 07-OCT-2002			
DEFINITION	UI-CF-DUI1-adp-o-08-0-UI.s1 UI-CF-DUI1 Homo sapiens cDNA clone			
	UI-CF-DUI1-adp-o-08-0-UI 3', mRNA sequence.			

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BU686683
BU686683.1 GI:23541826
EST.

Homo sapiens (human)
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 616)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL
MEDLINE
PubMed

Genome Res. 6 (9), 791-806 (1996)
97044477
8889548

COMMENT

Contact: McCray, PB
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA-Yes

FEATURES
source

Location/Qualifiers
1..616

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI1-adp-o-08-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-DUI1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GGCTGTAGGC.
TAG_Lib=UI-CF-DUI1
TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_SEQ=GGCTGTAGGC"

BASE COUNT
ORIGIN

195 a 104 c 86 g 231 t

Query Match
Best Local Similarity
Matches 592; Conservative

8.1%; Score 590.8; DB 13; Length 616;
99.7%; Pred. No. 1.9e-51;
0; Mismatches 2; Indels 0; Gaps 0;

QY

6637 ATGTTACCAGACAAACCCAGCTTTTGTCCCAAGGTGACTGTAACTCAGATGGAAGT 6696

Db

611 AGGTTACCAGACAAACCCAGCTTTTGTCCCAAGGTGACTGTAACTCAGATGGAAGT 552

QY

6697 GGGCTTTTATATAGGGTGGAGTCAACATCGTGTATCTTACTTAACAGCCCTTTGAA 6756

Db

551 GGGCTTTTATATAGGGTGGAGTCAACATCGTGTATCTTACTTAACAGCCCTTTGAA 492

QY

6757 TTTAACAAAACCTGGGAATCCATTAGGAACGGATTGTCATCATCACTCAATGAAGTCG 6816

RESULT 14
AG152578/C

LOCUS	AG152578	659 bp	DNA	linear	GSS 09-JAN-2002
DEFINITION	pan troglodytes dna				

accession: Pan troglodytes DNA, clone: RP43-016A13.TU, genomic survey
 sequence.
 ACCESSION AG152578
 VERSION AG152578.1 GI:16682256
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Pan.
 1
 REFERENCE Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of Library RPC1-43
 AUTHORS Unpublished
 TITLE
 JOURNAL

[illegible]

COMMENT

and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045; Japan (E-mail: chimbee@gsc.riken.go.jp, URL: <http://hgpc.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170).

Clones are derived from the chimpanzee BAC library RPCI-43. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: TJ
LIBRARY
Vector : pBACE3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI

```

FEATURES
  source
    Location/Qualifiers
      1. .659
        /organism="Pan troglodytes"
        K.site 2 : EGORI.

```

```

/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-016A13.TJ"
/sex="male"

```

BASE COUNT		ORIGIN	
207 a	122 c 96 g	232 t	2 others
/cell_type="lymphocytes"			
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"			

Query Match 7.6%; Score 556.4; DB 29; Length 659;
Best Local Similarity 94.5%; Pred. No. 6.1e-48;
Matches 598; Conservative 0; Mismatches 32; Indels 3; Gaps 2;

6652 ACCCAGCTTTTGTGCCAAGGTCACGTGTAACTCAGAAATCGAAAGTGGGCTTTATATAGG 6711

656 AACCCAGTTTGTCCCCAGGGTGGCTGTAACTCAAAATGGGAAGTGGCTTATATATAGG 597

6712 GTGTGAGTGAAGAACATGCTGTATGTTACTTACTTAACAGCCCTTTTGAATTTTAACAAAAAC-TG 6770
|||||
596 GTGTGAGTGAAGAACATGCTGTATGTTAGTACAGCCCTTTTGAATTTTAACAAAAACGTG 537

6771 GGAATCCATTAGGAACCGATTGCATCTACTGTAACAATAAGCTGCACTGCTGAAATTGT 6830
536 GAGATCCATTAGGAACCGATTGCATCTACTGTAACAATAAGCTGCACTGCTGAAATTGT 477

6831 ATTTTAGCTAATGAAAAGTGTTCGGACTAGTACTTAAAAATGTTCTAATGATAAAGT 6890
|||||
476 ATTTTAGCTAATGAAAAGTGTTCGGACTAGTACTTAAAAATGTTCTAATGATAAAGT 417

6891 TTTGAGTCAAAATACAAAGAAAAAATCTGCAATTCAGCCGAATTTTGTATATTTTA 6950

Mon Feb 2 09:20:11 2004

```

416 TTTGAGTCAAAATAGAAAAGAAAATCTGCATTCAGGCCGAAATTTTGTAATTTTA 357
6951 TTGCATTTAAATTTGCTATTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCG 7010
356 TGGCATTTAAATTTGCTATTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCA 297
7011 TGTACTTAAATTTGCTATTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCA 7070
296 TGTACTTAAATTTGCTATTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCA 237
7071 TATTTTTC--TGTGTATTAAGCTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCA 7128
236 TATTTTAAAGAGAGATGATATAGCTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCA 177
7129 ATACTACATTTTAAACATTCACATCCAAAGAGAGAGATGATATAGCTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCA 7188
176 ATACTACATTTTAAACATTCACATCCAAAGAGAGAGATGATATAGCTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCA 117
7189 TTTTAAATATTTAAATTTGCTATTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCA 7248
116 TTTTAAATATTTAAATTTGCTATTCTGTAATATTGGGAATCAAGTGGCTTATCATGTATATCA 7281
7249 TGTATATTTGCTGTAAGATTAAGATATGTA 7281
56 TGTATATTTGCTGTAAGATTAAGATATGTA 24

```

```

RESULT 15
AG120959/c 673 bp DNA linear GSS 04-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-129N20.F, genomic survey sequence.
DEFINITION AG120959
ACCESSION AG120959
VERSION AG120959.1 GI:16650124
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 673)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. .673
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-129N20.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
238 a 127 c 120 g 188 t

```

COMMENT

```

Query Match 7.5%; Score 549.8; DB 29; Length 673;
Best Local Similarity 95.8%; Pred. No. 2.8e-47; Indels 4; Gaps 3;
Matches 597; Conservative 0; Mismatches 22;
QY 5561 TTTTAAAAATAGAGCTTGGTCAACCTCTATACACATATTACAAAGATATAGACCTTTT 5620
672 TGTAAAAATAAAGCTTGGTCAACTCATATCCATATCCACATATTCCAGATATAGCCCTTTC 613
5621 AAAATCAATCTAAACCTTTACAGAACTTTCTTATAGGTTATGCCCTTTTATTTTAAAGACT 5680
612 AAAATGATTTCTAAACCTTTTACAAAACCTTTCTTATAGGTTATGCCCTTTTATTTTAAAGACT 553
5681 TATTATAAATCAAGTCCCATTTAGATGATATATATAGTCTTGTAGGCTTTGATATATAGCTTTG 5740
552 TATTATAAATCAAGTCCCATTTAGATGATATATATAGTCTTGTAGGCTTTTAAATATATAGCTTTG 493
5741 TGTACAAAATGTTAGTATGTTATTTTAAACAGGTACATTTTATACA--GTGTTTTCTTTATC 5798
492 TGTACAAAATGTTAGTATGTTATTTTAAACAGGTACATTTTATACA--GTGTTTTCTTTATC 433
5799 AATTTCGTATATGTCACAGAAATCAGTGTGTCTTTTTCATAAGGTTTTTCAATGCTTTAT 5858
432 AATTTCGTATATGTCACAGAAATCAGTGTGTCTTTTTCATAAGGTTTTTCAATGCTTTAT 373
5859 TTTTTCACAAAGTTTACGTTCTCAAGACACACTGTCTTCCAGTACGTAAGTTAAAAA 5918
372 TTTTTCACAAAGTTTACGTTCTCAAGACACACTGTCTTCCAGTACGTAAGTTAAAAA 313
5919 TACAGTTTACCAAGTTGCTTCTAGGCTACTGAGATCCATGTGACATTTGGAGGAGATCT 5978
312 TACAGTTTACCAAGTTGCTTCTAGGCTACTGAGATCCATGTGACATTTGGAGGAGATCT 253
5979 TTTTAAATGTTTATGATTTTCTGTCATTAGCAATGGCTGTTTGTAGTTTGTGTTAAATGTGTGC 6038
252 TTTTAAATGTTTATGATTTTCTGTCATTAGCAATGGCTGTTTGTAGTTTGTGTTAAATGTGTGC 193
6039 CTAAGTTGAATTTGTTGTTTTTCTTCTACACTGTGTACAGGCTTGTCTACACACAGA 6098
192 CTAAGTTGAATTTGTTGTTTTTCTTCTACACTGTGTACAGGCTTGTCTACACACAGA 133
6099 TAAGTCTGTTGTGATCAGATAGATCTACATAGTTTGTGAGTTTGTGTTAAATGTGTTAAATGTG 6157
132 T-AGTCTGTTGTGATCAGATAGATCTACATAGTTTGTGAGTTTGTGTTAAATGTGTTAAATGTG 74
6158 TAGGAGGCTTCCCGGAGCTCGTA 6180
73 TAGGAGGCTTCCCGGAGCTCGTA 51

```

Search completed: February 1, 2004, 00:08:15
Job time : 13523 secs

FEATURES
source

BASE COUNT
ORIGIN